# Essays on organizational economics and the heterogeneity of firm performance

by

Sarah Elizabeth Venables

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013





Author ..........................................................................
Department of Economics
April 18th, 2013

Certified by ......................................................................
Robert Gibbons
Professor
Thesis Supervisor

Certified by ......................................................................
Glenn Ellison
Professor
Thesis Supervisor

Accepted by ......................................................................
Michael Greenstone
Chairman, Department Committee on Graduate Theses

## Essays on organizational economics and the heterogeneity of firm performance

by

Sarah Elizabeth Venables

Submitted to the Department of Economics
on April 18th, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

### Abstract

The first chapter presents a model in which a principal requires an agent to work on multiple tasks, which require varying levels of effort. Initially the appropriate actions are unknown to the principal and so he cannot monitor the agent's effort; however he may succeed in learning the game over time. I provide conditions that ensure full learning and cooperation on all tasks. If these conditions are not met, then the performance of the relationship may depend critically on the order in which tasks are first required and on whether early attempts are successful. We may therefore see variation in long-run performance amongst partnerships facing an identical environment.

The second chapter considers the role that mission statements made by firms may play in influencing workers' expectations, and whether these mission statements may thereby constitute informative cheap-talk. I consider a setting in which firms send costless mission statements at the start of the game, then with positive probability have an opportunity to send a costly signal of their type at a later stage. I show that this can generate informative cheap-talk in environments in which, absent the prospect of a costly signal, cheap-talk messages would otherwise be uninformative. However firms will face a welfare trade-off between informative communication and costly signaling, and moreover may be constrained in their ability to adapt to changing circumstances.

In the third chapter I present a model of relational contracting in which the principal's type determines the probability that effort generates high output, and where this type is private information. In this setting the bonus payments will be driven by the agent's beliefs, and by the principal's desire to signal her type. I show that we will expect to see higher bonus payments following a run of poor output, given the need to restore the agent's confidence. I then consider an extension in which the agent receives a signal about the performance of the principal's competitors, and so the principal will be under more pressure to "match the market". I argue that this model may explain the dynamics of bonus payments following recessions.

Thesis Supervisor: Robert Gibbons
Title: Professor

Thesis Supervisor: Glenn Ellison
Title: Professor

## Acknowledgments

I would like to thank my advisors Robert Gibbons and Glenn Ellison for their support over the past few years. I have benefitted greatly from their advice and encouragement and am grateful to have had the opportunity to work with them.

I would also like to thank my classmates and colleagues who have been such an integral part of my experience at MIT. In particular, I would like to thank Eliza Forsythe, Daniel Barron, Michael Powell, Hongyi Li for always being open to discussing ideas. I would also like to thank the faculty and students who attended the Organizational Economics lunch over the years for their feedback and suggestions.

Lastly, I am tremendously grateful to Jennifer Peck and Xiao Yu Wang who made Cambridge feel like home for the past five years. I will always appreciate the friendship, conversation and scotch that they so generously supplied.

# Contents

1 **Learning to Cooperate: Building Relational Contracts in Environments with Imperfect Monitoring** **9**

1.1 Introduction . . . 9
1.2 Related Literature . . . 11
1.3 Model . . . 13
1.3.1 Benchmark Cases . . . 14
1.4 Equilibrium in the game with private information . . . 16
1.4.1 First-Best . . . 17
1.4.2 Second-Best . . . 19
1.4.3 Equilibria with "probation" . . . 25
1.5 Example . . . 27
1.5.1 Example with two tasks . . . 28
1.5.2 Example with three tasks . . . 32
1.5.3 Path Dependence . . . 35
1.6 Persistent Performance Differences . . . 37
1.7 Conclusion . . . 41
1.8 Appendix . . . 42
1.8.1 Proof of Proposition 3 . . . 42
1.8.2 Proof of Proposition 4 . . . 43
1.8.3 Details from the examples in section 5 . . . 45

2 **Walking The Talk: Mission Statements as Cheap Talk** **55**

2.1 Introduction . . . 55
2.2 Related Literature . . . 58
2.3 Model . . . 59
2.4 Equilibrium . . . 61
2.4.1 Increasing the probability that the signaling game is played . . . 64
2.4.2 Equilibria in which more than two messages are sent . . . 67
2.4.3 Increasing the cost of signaling . . . 69
2.5 Example . . . 71

2.5.1 Welfare . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
2.6 Adaptation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
2.7 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
2.8 Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81
2.8.1 Proof of Proposition 1 . . . . . . . . . . . . . . . . . . . . . . . . . . 81
2.8.2 Proof of Lemma 1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
2.8.3 Proof of Lemma 2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

**3 Matching the Market: Relational Contracts and the Dynamics of Bonus Payments** **85**
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85
3.2 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88
3.3 Equilibrium . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
3.3.1 Equilibrium when the principal's type is private information . . . . . . 91
3.3.2 Separating equilibrium in the game with private information . . . . . . 92
3.3.3 Equilibrium in which both types pool in the short-run . . . . . . . . . . 97
3.3.4 Optimal equilibrium for the high-type . . . . . . . . . . . . . . . . . . . 101
3.4 Matching the Market . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
3.4.1 Inducing separation through payment of the bonus. . . . . . . . . . . . 105
3.4.2 Equilibrium in which types pool in the short run . . . . . . . . . . . . . 106
3.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 109

# Chapter 1

# Learning to Cooperate: Building Relational Contracts in Environments with Imperfect Monitoring

## 1.1 Introduction

There is an extensive literature in economics showing that even in the absence of formal contracting and incentives it may be possible for parties to sustain a cooperative outcome in an infinitely repeated game through a relational contract. However this implicitly assumes that it is clear to both parties what constitutes cooperation. Often, however, cooperative behaviour may be hard to define. This may be because workers are unfamiliar with a particular environment or task, and need time to experiment before they can provide consummate effort, or because workers are unwilling to invest costly effort in a project until they are sure that it is necessary and worthwhile. In both these cases we may imagine that actions could be misinterpreted, potentially leading to the breakdown of a relationship. However if a partnership has survived these initial hurdles then as time goes on and those involved have a better understanding of the others' behaviour it should become easier to maintain a cooperative relationship.

In this paper I propose a model of relational contracting in an infinitely repeated game in which a risk-neutral principal and agent must collaborate on a finite number of tasks, one of which will be required in each period. These tasks will differ in the level of effort that they require from the agent: some tasks require only a minimal level of effort, which I will assume is costless, others require the agent to take a more costly action. If the agent takes the appropriate action, then the task will be completed successfully with positive probability; if she takes the incorrect action then there is no chance of success. The cost of effort and the probability of success are such that it is optimal for the agent to supply effort on all tasks. I will assume that the principal can observe the agent's actions, and whether or not the task is a success, but does not know the mapping from tasks to correct actions. This, combined with the fact that taking the correct action cannot guarantee a success, means that the principal cannot perfectly monitor the agent's behaviour. The paper proposes an environment without formal contracting and with limited liability on the part of the agent, and so the means through which the principal can provide the correct incentives to the agent are limited to the promise of non-negative monetary transfers and the threat of terminating the relationship.

As an example of a setting in which we may expect there to be asymmetric information regarding what constitutes the appropriate level of effort we could consider the R&D process. Different stages of a project may require different levels of effort and whilst at times the agent would be expected to be very busy, at other times she may be waiting for results and thus appear to be supplying less effort. If the principal is unfamiliar with the particular research process then he may be unable to tell whether the agent is behaving correctly, given what the project requires, or whether she is shirking. Moreover, the inherent uncertainty involved in R& D projects means that the principal cannot necessarily attribute the failure of a project to the agent's actions. However as the principal gains in experience and has observed the outcomes of earlier projects, he is able to monitor the agent's behaviour more closely. We could also think about modelling a more basic manufacturing procedure, such as a production line, in this manner. Whilst quotidian tasks may require little input from the agent, at times she may be required to put in extra effort in order to resolve a problem or increase the efficiency of the process. Although the principal can observe when this extra effort is supplied, if he cannot identify the situations in which the agent should be supplying a higher level of effort then he cannot identify whether failures are due to the agent shirking, or to chance. Once again, as the principal observes more of the agent's behaviour in different situations, and the outcomes of her actions, he is better able to monitor her effort in the future.

To analyse this model, I will begin by considering the full-information benchmark in which the actions required on each task are common knowledge, and show that in this setting cooperation on all tasks can be sustained if the discount factor is sufficiently high. Failing this, it may be possible to achieve cooperation on only the costless tasks; it will never be optimal to work only on a subset of the costly tasks. I then turn to the case with imperfect monitoring and I show that in this setting there will be values of the discount factor $\delta$ for which full cooperation on all tasks is possible in the game with full information, but cannot be achieved in the game with private information. In particular, I show that there are three potential causes of inefficiency or relationship breakdown: (i) following certain histories the parties' beliefs about the future surplus of the relationship may drop to a level at which they no longer wish to continue the game, even if there is in fact sufficient surplus to sustain cooperation; (ii) the agent may attempt to deceive the principal about the true value of the relationship in order to extract more surplus, making the principal skeptical of her behaviour even if she is behaving truthfully; and (iii) the principal may be unable to elicit effort on a particular task without using the threat of termination. Any one of these factors could generate an outcome in which the principal and the agent either terminate the relationship entirely, or work only on a subset of tasks. Moreover, I show that the eventual outcome is highly dependent not only on the realised costs of tasks, but on the order in which they are required and whether early attempts are successful. I go on to show that relaxing the restrictions on the equilibrium to allow for a "probationary period" during which the agent is not expected to provide costly effort will alleviate some of these problems and reduce the discount factor necessary to induce cooperation, however it will not return the relationship to the full-information level of efficiency.

Given that eventual outcomes are highly dependent on initial draws in this setting, this model may provide a possible explanation for the differences we see in performance amongst firms that operate in the same industries. These large differences in performance have been shown to exist even in narrowly defined industries and amongst firms producing homogeneous goods (Foster et al. 2008, Syverson 2004). Moreover, these differences have been shown to be highly persistent over time (Melitz 2003, Baily et al., 1992, Foster et al., 2006). This paper suggests that these differences may arise if some firms draw costly tasks at the start of the game, before they are sufficiently confident of the value of the relationship. These differences in eventual outcomes could also occur if some firms are unsuccessful on early tasks, whereas others are successful. These gaps in performance may occur even if the set of tasks (and the costs of those tasks) is identical across firms. The paper goes on to ask whether the gaps in performance can be closed if firms can observe something about the performance of other firms in the industry. That is, is it easier to sustain cooperation in this game if the parties involved can infer something about the total value of the relationship from the behaviour of others?

The paper proceeds as follows. The next section sets out the related literature. In section 3 I describe the model, and analyse two benchmark cases against which I evaluate the the performance of the partnership. Section 4 provides the main results of the paper, which are then illustrated through two examples in section 5. Section 6 returns to the question of persistent performance differences and considers whether allowing firms to observe a coarse indicator of other, more successful, firms' behaviour can enable them to close the gap in performance. Section 7 concludes.

## 1.2 Related Literature

This paper builds on the literature on relational contracts in its analysis of the provision of incentives in a repeated game in which output is not contractible. In the later stages of the game, once the mapping from actions to output on each task is common knowledge, the problem for the firm is one of incentive provision, similar to that in Levin (2003). The earlier stages of the game, however, share more characteristics with the work on imperfect public monitoring (Fudenberg-Levine-Maskin 1994, Green and Porter 1984) in which players only observe an imperfect signal of others' actions and use this imperfect signal to infer deviations, meaning that with positive probability punishments may occur on the equilibrium path. In this setting the principal observes actions perfectly, but does not initially know the mapping from actions to output; imperfect monitoring comes from the fact that the principal cannot distinguish whether a failure was due to the agent shirking or to chance.

This paper can also be related to the literature on establishing collusion across multiple markets, as in Bernheim and Whinston (1990). As in that paper, in this setting slackness in incentive constraints in certain markets, or on certain tasks, can be used to enforce collusive behaviour in other markets or tasks. This may make it possible to sustain cooperative behaviour for values of the discount factor for which it would not be possible to do so were markets treated independently. This paper adds the problem of imperfect monitoring, and shows that the need to provide incentives on other tasks may in fact hamper the principal's ability to enforce cooperation on a new task.

Another paper that deals with the problem of establishing a cooperative relationship in a setting in which there is asymmetric information about what such cooperation entails is Chassang (2010). That paper also has the result that inefficient termination may occur on the equilibrium path, and that different partnerships may end up facing different outcomes as a result of the early history of the game. This paper differs from Chassang's in that it allows for the principal to promise variable transfers based on output. Moreover this paper also incorporates *a priori* uncertainty about the expected surplus from the relationship, and so the asymmetry of information between the principal and the agent extends to the total value of the relationship, as well as the actions required in any particular period.

Other papers have also explored cooperation games in settings in which there is uncertainty amongst agents regarding the appropriate actions required to generate a success. For example Crawford and Haller (1990) consider an environment in which players know the mapping from actions to payoffs but lack any means of communicating their strategy to the other players, meaning that experimentation is required at the start of the game, but cooperation becomes easier as the game proceeds and parties can communicate by referring to the history of the game. Blume and Franco (2007) analyse a somewhat similar game in which the mapping from actions to payoffs is initially completely unknown, and in addition players lack the means to describe their actions or strategies to other players; optimal experimentation in this context requires that each player switches actions with some probability in each period. More closely related to this paper, Blume, Franco and Heidhues (2012) extend the previous model to allow each player to have some private signal about the correct action to be taken, the optimal strategy must now incorporate this information. In this paper the interests of the players are fully aligned, but they are unable to commicate their informative. This paper differs from these mentioned since it adds an incentive problem:

whilst one player has (initially incomplete) information about the mapping from actions to payoffs, this player may have an incentive to conceal that information. Later sections of this paper consider whether more formal rules concerning experimentation can alleviate the incentive problem.

## 1.3 Model

A principal (he) and an agent (she) are required to work together on a series of tasks. Let the set of tasks be denoted by $\mathscr{T} = \{1, ..., n\}$. In each period one task is drawn at random. The probability that task $i$ is drawn in any period is $\alpha_i$, where $\sum_i \alpha_i = 1$. These probabilities are independent and identical over time. Each task requires that the agent take an appropriate action. Assume for simplicity that the agent has two available actions: $a \in \mathscr{A} = \{a_0, a_c\}$, where $a_0$ denotes the costless or default action, and $a_c$ denotes the costly action, with $c(a_c) = c$. If the agent takes the correct action, then the task will be successful with probability $\eta < 1$; otherwise the task will fail. Let the probability that the task requires the costly action be $p$. If the task is successful then value 1 accrues to the principal. Assume that $\eta > c$, and so total expected surplus is maximised by having the agent take the correct action on all tasks. Let $C \in \mathscr{C} = \{0, c\}^{\mathscr{T}}$ denote a particular *cost profile*, that is, $C = (a_1, ..., a_n)$ indicates the appropriate action on each task. Further, let $\gamma$ represent the total fraction of tasks which are costly, $\gamma = \sum_i \alpha_i \mathbb{I}\{a_i = a_c\}$.

The information and contracting environment is as follows: in each period $t$ both the principal and the agent observe the task drawn, the action taken by the agent $a_t$, and output $y_t \in \{0, 1\}$. However neither the action nor output are contractible. Assume that when task $i$ is first drawn, the agent learns the appropriate action for that task $a_i$, but that this is not observed by the principal. If a success is observed on task $i$, then, knowing the action taken by the agent, the principal learns $a_i$; however if a task is unsuccessful then the principal cannot distinguish between the agent having taken the incorrect action, and the agent taking the correct action but being unsuccessful. There is no formal contract available to the principal, but he may promise a bonus payment $\tau_i$ to the agent if the task is successful. I will assume that the agent has a limited liability constraint and so transfers must be non-negative.

The relationship thus faces two potential issues. Firstly, the principal and the agent are unable to contract on actions or output, therefore any bonus payments can only be sustained through the threat of termination of the relationship. This bounds the amount that it is credible for the principal to pay. Secondly, the principal cannot fully monitor the agent's actions, since he does not initially know which actions are appropriate.

The relationship is modelled as an infinitely repeated game. In each period, the following stage game occurs:

1. The principal offers the agent a relational contract, which the agent may accept or

reject. If the agent rejects, both players receive their outside option.

2. A task $i$ is drawn, and observed by both the principal and the agent.

3. If the task $i$ has not previously been encountered, the agent learns the appropriate action $a(i) \in \{a_0, a_c\}$.

4. The agent chooses an action $a \in \{a_0, a_c\}$ at a cost $c(a)$. This action is observed by both the principal and the agent.

5. Output $y \in \{0, 1\}$ is realised, and observed by both parties.

6. The principal may make a transfer $\tau$ to the agent.

7. Payoffs are received.

Both the principal and the agent are assumed to be risk neutral. The agent's expected utility in each period is therefore her expected transfer, net of any costs of effort. The principal's profit is his expected surplus net of transfers paid to the agent. Assume that the principal and the agent share a common discount factor $\delta$. Let the agent's outside option have a net present value of $U_0$, and the principal's outside option have a net present value of $\Pi_0$; assume that $-c + \eta \geq (1 - \delta)(\Pi_0 + U_0)$ and therefore it is efficient for the principal and the agent to work together even if all tasks are costly.

### 1.3.1 Benchmark Cases

Before proceeding to analyse the full model with private information and no formal contracts, it is worth relaxing these assumptions in order to create a benchmark against which to evaluate the efficiency of the equilibrium in the game with private information. I will consider both the benchmark in which output is perfectly contractible, and then the case in which output is non-contractible, but the correct action on each task is common knowledge.

The first benchmark case is that in which output is fully contractible. Since both are risk-neutral, the principal can use an optimal incentive contract to induce the agent to take the correct actions, even though the principal cannot contract directly on the action. For example, the first-best can be achieved by having the principal pay a transfer $\tau_{FB}$ following any success, where $\tau \geq \max\{c/\eta, (\gamma c + (1 - \delta)U_0)/\eta\}$. Setting $\tau_{FB} \geq c/\eta$ ensures that the agent will take the costly action when appropriate, whilst $\tau_{FB} \geq (\gamma c + (1 - \delta)U_0)/\eta$ ensures that the agent is willing to remain in the relationship.

The other benchmark that may be considered is the game with full information, that is, the game in which the principal knows $C$ and so can fully monitor the agent's actions. Although the principal and the agent cannot formally contract in this setting, for a sufficiently high discount value $\delta$ they may be able to ensure cooperation through a relational contract. In this case, following Levin (2004), a stationary relational contract can do as well as any other contract. Let the contract take the following form: the agent will take the appropriate action in every period, and the principal will pay a bonus $\tau_i$ if the task is a success. In this setting with perfect monitoring there need not be any punishment on the equilibrium path, and therefore the optimal relational contract can maximise punishments for deviation by having the relationship be terminated whenever either the agent takes the incorrect action, or the principal fails to pay the appropriate transfer.

The following conditions are necessary to sustain the correct actions: the agent must be willing to take the costly action whenever appropriate, and the principal must be willing to pay the specified transfer. The former condition requires that:

$$-c_i + \eta\tau_i + \delta EU \geq \delta U_0, \tag{1.1}$$

and the latter that:

$$-\tau_i + \delta E\Pi \geq \delta Pi_0, \tag{1.2}$$

for all $i$.

**Proposition 1.** *The maximal payoff that can be generated by a relational contract can be achieved by a contract that sets $\tau_i = \tau_j$ for all $i, j$.*

*Proof.* Combining the two constraints (1.1) and (1.2) above, we know that the relationship can be sustained on a given task $i$ only if

$$\delta \geq \frac{c}{\delta(EU - U_0) + \eta(E\Pi - \Pi_0)}.$$

Since $\eta < 1$, it is therefore optimal to shift surplus from the principal to the agent, subject to the constraint that $-\max_i\{\tau_i\} + \delta E\Pi \geq \delta\Pi_0$. This can be achieved by having the principal's constraint bind for all $i$, implying that $\tau_i = \tau_j$, for all $i, j$. □

The intuition for this result is that if the principal's payment constraint doesn't bind for all $\tau_i$, that is, if there is slack in this constraint for certain $\tau_i$, then this should be transferred to the agent in order to increase her continuation value. Although doing so will reduce the principal's future surplus, and hence the maximum transfer that he is willing to pay, for $\eta < 1$ this reduction in a particular transfer is less than the gain to the agent through future surplus, and this will increase the set of values of $\delta$ for which the agent can be induced to supply costly effort.

Therefore the optimal relational contract under full information has the principal pay a transfer $\tau$ to the agent following any success, even on a costless tasks. Letting $\gamma$ denote the fraction of tasks that require the agent to take the costly action, the constraints above become:

$$-c + \eta\tau + \frac{\delta}{1-\delta}(-\gamma c + \eta\tau) \geq \delta U_0,$$

$$-\tau + \frac{\delta}{1-\delta}\eta(1-\tau) \geq \delta\Pi_0.$$

The relationship can therefore be sustained under full information whenever there exists $\tau$ such that

$$\frac{c(1-\delta+\delta\gamma)+U_0\delta(1-\delta)}{\eta} \leq \tau \leq \frac{\delta\eta - \Pi_0\delta(1-\delta)}{1-\delta+\delta\eta}. \tag{1.3}$$

Therefore, for a given fraction of costly tasks $\gamma$, the relationship can be sustained under full information for all $\delta \geq \delta^*_\gamma$, where $\delta^*_\gamma$ is found by having the two constraints above hold with equality. These cutoffs are increasing in $\gamma$: a higher fraction of costly tasks decreases the total surplus from the relationship, making it more difficult to elicit effort on these tasks.

For $\delta < \delta^*_\gamma$ cooperation on all tasks will not be possible. In this case the principal and the agent may be able to achieve a partial level of success by agreeing to work only on the costless tasks. However they cannot do better by only working on a subset of the costly tasks: given that $\eta > c$, all tasks generate positive surplus in expectation, so eliminating certain costly tasks only reduces total surplus, it does not make it easier to provide the correct incentives on the remaining costly tasks.

The agent can be persuaded to remain in the relationship and work on the costless tasks if the expected surplus from doing so exceeds her outside option:

$$\eta\tau + \frac{\delta}{1-\delta}(1-\gamma)\eta\tau \geq \delta U_0,$$

where the promised transfers must be incentive compatible for the principal:

$$-\tau + \frac{\delta}{1-\delta}(1-\gamma)\eta(1-\tau) \geq \delta\Pi_0.$$

Combining these constraints, cooperation can be sustained if there is some $\tau$ that satisfies:

$$\frac{1-\delta}{(1-\gamma)\eta}U_0 \leq \tau \leq \frac{\delta(1-\gamma)\eta - \delta(1-\delta)\Pi_0}{1-\delta+\delta(1-\gamma)\eta} \tag{1.4}$$

which yields a series of cutoffs $\{\delta^0_\gamma\}$, where $\delta^0_\gamma < \delta^*_\gamma$ for all $\gamma$, for which cooperation can be sustained only on the fraction $1-\gamma$ of costless tasks.

## 1.4 Equilibrium in the game with private information

I now turn to the game with private information, and show that for sufficiently high $\delta$ it is possible to obtain the same level of performance as in the game with full information. I then consider lower values of the discount factor, and show that there exist parameter values for which optimal performance is possible in the game with full information, but the relationship may be terminated inefficiently or operate at a lower level of output when the correct actions are not common knowledge.

In the game with private information the principal does not initially know $C$, but may learn it over time, after which cooperation can be sustained as in the full information case above. In the interim, the principal cannot perfectly monitor the agent's behaviour but can use transfers and the threat of termination to induce the agent to cooperate. Since the principal cannot distinguish between the agent correctly taking the costless action but being unsuccessful, and the agent shirking, she may need to punish the agent whenever he takes the costless action without success. The principal can either punish the agent by reducing expected surplus through the reduction of transfers on other tasks in the futures, or through the threat of terminating the relationship altogether.

I look for a Perfect Bayesian Equilibrium of this game, and will first consider an equilibrium which is also subject to the following conditions: (i) the agent always takes the correct action $a(i) = C(i)$; (ii) if a success is realised on task $i$, the principal pays a transfer $\tau_i(h^t)$; (iii) if at any time $s$ task $i$ is drawn and the agent took action $a_s(i)$, then if at any time $t > s$ task $i$ is drawn and the agent takes an action $a_t(i) \neq a_s(i)$ then the relationship is terminated; and (iv) if the principal fails to pay the specified transfer following a success then the relationship is terminated. That is, I will first look for an equilibrium in which the agent takes the correct actions from the start of the game, and the relationship is terminated if she ever changes her action on a particular task. I later relax this assumption to consider an equilibrium in which the agent is allowed to shirk on costly tasks at the start of the game while she obtains information about $C$.

### 1.4.1 First-Best

For sufficiently high $\delta$, the principal and the agent may be able to sustain correct behaviour on all tasks without any risk of the relationship breaking down on the equilibrium path. Since neither the principal nor the agent know how many tasks will require the costly action there is initial uncertainty about the value of the relationship, and so for any $\delta < \delta_1^*$ there is a positive probability that $C$ will be such that full cooperation cannot be sustained even under full information. It is only for $\delta \geq \delta_1^*$ that the principal and the agent can be sure that the relationship will succeed under full information, for any draw of tasks. In this case there is sufficient surplus to ensure cooperation on all tasks in the game with private information, and there will be no risk of the relationship breaking down.

**Proposition 2.** *For $\delta > \delta_1^*$ there exists $\{\tau_i\}_{i \in \mathcal{I}}$ such that agent supplies the correct level of effort $a_t(i) = a_i$ on all tasks $i$.*

*Proof.* This result can be shown by constructing an equilibrium which satisfies the following properties: (i) the agent always takes the correct action; (ii) the principal pays a transfer $\tau_c = \frac{c+U_0\delta(1-\delta)}{\eta}$ if the agent takes a costly action and is successful, and a transfer $\tau_0 = \frac{\delta U_0}{\eta}$ if the agent takes a costless action and is successful; and (iii) the agent quits if the principal fails to pay the transfer when appropriate.

In order to ensure that the agent takes the correct action $a_i^*$ on task $i$, then we require, for any prior history $h^t$, that

$$-c(a_i^*)+\eta\tau_c+\delta E[U(h^t,a_t(i))|a_t(i)=a_i] \geq \max\{0+\delta E[U(h^t,a_t(i))|a_t(i)\neq a_i],\delta U_0\}, \quad (1.5)$$

where $E[U(h^t,a_t(i))|a_t(i)=a_i]$ is the agent's expected continuation value from taking action $a_t$ on task $i$, conditional on the correct action being $a_i$. Since $-c+\eta\tau_c \geq 0$ the expected payoff from correctly taking the costly action on task $i$ in the current period exceeds the payoff from deviating. Moreover, the principal can ensure that $E[U(h^t,a_t(i)),a_t(i)=a_i] \geq E[U(h^t,a_t(i))|a_t(i)\neq a_i]$ by making the agent's expected payoffs on all tasks $j\neq i$ independent of the outcome of task $i$. Therefore this transfer $\tau_c$ is sufficient to induce the agent to take the costly action when appropriate. If we turn to the tasks for which the costless action is appropriate, we need to ensure that the agent does not quit the relationship:

$$\eta\tau_0+\delta E[U(h^t,a_t(i))|a_t(i)=a_i] \geq \delta U_0.$$

For the specified transfers, it must be the case that $E[U(h^t,a_t(i))|a_t(i)=a_i] \geq 0$, and therefore the transfer $\tau_0=\frac{\delta U_0}{\eta}$ is sufficient to ensure that the agent remains in the relationship.

It remains to check that the principal is willing to pay these transfers whenever appropriate. This requires that

$$-\max\{\tau_c,\tau_0\}+\delta E\Pi(h^t) \geq \delta\Pi_0.$$

We know that $\tau_c > \tau_0$, and therefore that $E\Pi \leq \frac{1}{1-\delta}\eta(1-\tau_c)$. A sufficient condition for the principal to be willing to pay these transfers is therefore that:

$$-\tau_c+\frac{\delta}{1-\delta}\eta(1-\tau_c) \geq \delta\Pi_0$$

where $\tau_c=\frac{c+U_0\delta(1-\delta)}{\eta}$. But from the full information case, we know that $\delta_1^*$ is the discount factor for which this expression will hold with equality, and therefore full cooperation can be assured on any draw of tasks for any $\delta \geq \delta_1^*$. □

The intuition for this result is that if the discount factor is sufficiently high that the principal can commit to a transfer that will induce effort if *all* tasks are costly, then that same transfer can be used on a task-by-task basis to induce the agent to take the correct action. There is therefore no need to use the surplus generated on the costless tasks to cross-subsidise the agent on the costly tasks. Whilst the agent still faces a temptation to shirk since the principal cannot identify this as incorrect behaviour, shirking will never generate a success whereas the expected transfer from taking the (correct) costly action exceeds the immediate cost.

### 1.4.2 Second-Best

In this section I consider values of $\delta$ for which it is not possible to sustain the relationship if all tasks are costly, and so there is uncertainty about whether the relationship will survive. Moreover, it may be necessary to cross-subsidize tasks: that is, to use the surplus generated on costless tasks to provide the incentives necessary to induce effort on costly tasks. I will begin by considering the subgame in which only one task remains unknown to the principal, and show that it may not be possible to provide the necessary incentives on this task without using the threat of inefficient termination. I then extend this argument to the general case, and show that the uncertainty about the surplus generated will add another source of inefficiency

It is worth noting that for any $\delta < \delta_1^*$ there will always be some risk of *efficient* termination on the equilibrium path. That is, if $C$ is such that cooperation cannot be sustained under full information, then the relationship must also break down in the game with private information. However there is also a positive probability of inefficient termination on the equilibrium path. There are three possible sources of this inefficiency. The first source of inefficiency somes from the uncertainty about $C$, and hence about the true value of the relationship. Even if $C$ is such that cooperation could be maintained under full information, following certain histories the parties' beliefs about the future surplus may drop to the point at which further cooperation on costly tasks is not possible. This in turn affects the transfers that the principal can commit to pay, since the risk of breakdown reduces the expected surplus from the relationship, making it harder to elicit cooperation from the agent.

Secondly, in addition to this symmetric *ex ante* uncertainty, there may also be asymmetric uncertainty about future surplus. On drawing a task for the first time the agent learns the action required to complete the task and updates her beliefs about the value of the relationship. If an equilibrium exists in which the agent always takes the correct action (or quits), then at the end of each period the principal also updates his belief about the value of the relationship. However following certain histories the agent may have an additional incentive to deviate in order to conceal information about the value of the relationship, and thereby extract further surplus from the principal. Whilst this deception cannot persist in the long term, it may be profitable for the agent to attempt to deceive the principal in this way in the short-run (if, for example, the relationship would be terminated as soon as the principal learned the true value). Moreover, even if the agent is not deceiving the principal, if her actions are unsuccessful then the principal may still be skeptical about the true value of the relationship: this will in turn reduce the transfers to which he can commit, making cooperation more difficult.

The third issue is that of incentive provision. Given that the principal does not know the appropriate action for each task and so cannot perfectly monitor effort, the threat of termination may be required to induce the agent to cooperate. Since $\eta < 1$ and so the agent may in equilibrium take the correct (costless) action and fail to achieve a success, this will lead to inefficient termination on the equilibrium path with positive probability.

These three sources of inefficiency mean that the first-best will often be unattainable, in the sense that for any draw $C$ of costs across tasks, such that a fraction $\gamma$ require a costly action, there will exist values of $\delta$ for which cooperation on all tasks is possible in the game with full information, but there will be a positive probability of termination in the game with private information. Whether or not this deterioration in the relationship occurs may depend critically on the order in which tasks are drawn. In particular, if costly tasks are required at the start of the game then this reduces the expected surplus from the relationship, and the ability of the principal to commit to the transfers necessary to support full cooperation.

**Game in which only one task remains unknown.**

To state these results more formally, I will first focus on the case in which there is just one task $i$, occuring with probability $\alpha_i$, for which the correct action remains unknown. Assume that at period $t$, following history $h^t$, all tasks except task $i$ have been attempted successfully and so the correct actions on these tasks are common knowledge. Assume that a fraction $\gamma_{-i}$ of these tasks are known to require the costly action. Therefore if $a_i = a_c$, the total share of tasks that require the costly action is $\gamma_{-i}(1-\alpha_i)+\alpha_i$; otherwise, if $a_i = a_0$ a fraction $\gamma_{-i}(1-\alpha_i)$ of tasks require the costly action. We can then define the values of $\delta$ for which full cooperation would be possible in the game without private information, in either scenario: if the last task is costly, then we require $\delta \geq \delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}$, if costless, then we need $\delta \geq \delta^*_{\gamma_{-i}(1-\alpha_i)}$, where $\delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i} > \delta^*_{\gamma_{-i}(1-\alpha_i)}$.

There are two cases to consider: firstly, the case in which the tasks already drawn are such that there is sufficient surplus to sustain cooperation on either a costly or a costless task; and secondly, the case in which the known surplus from the relationship is near the cutoff such that task $i$ being costless would allow the principal and the agent to reach cooperation on all tasks, but a costly task will make cooperation on *any* costly task impossible. In the former case, where $\delta \geq \delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}$, it is common knowledge that there is sufficient surplus to maintain the relationship, but nonetheless the principal wishes to ensure that the agent takes the costly action if required. In the latter case, if $\delta \in [\delta^*_{\gamma_{-i}(1-\alpha_i)}, \delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i})$ there is no way of inducing the agent to take the costly task, but the agent may have an incentive to misrepresent the value of the relationship to the principal in order to obtain some surplus in the short term. In either case we may see the relationship be terminated with positive probability on the equilibrium path if the agent takes the costless action without success.

**Proposition 3.** *There exist values of $\alpha_i$, $\gamma_{-i}$ such that for $\delta \in [\delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}, \tilde{\delta}_{\gamma_{-i}(1-\alpha_i)+\alpha_i})$ there will be inefficiently low performance in equilibrium even though it is common knowledge that there is sufficient surplus to sustain the relationship at the optimal level.*

This result states that there exist values of $\delta \geq \delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}$ for which it is either impossible to induce the agent to take a costly action when appropriate, or doing so requires the threat of inefficient termination on the equilibrium path. This occurs for values of $\delta$ for which it is common knowledge that there is sufficient surplus to induce the agent to take a costly action on task $i$. Therefore the problem is not that the principal cannot commit to the necessary transfer once he knows that task $i$ requires the costly action, but that he doesn't know whether it requires this action.

A full proof is in the appendix, but the intuition is that the principal's ability to provide incentives through the promise of future surplus is limited by the fact that some of this surplus has already been pledged to the agent in order to provide incentives on earlier tasks. In order for the agent to be induced to undertake the costly action, it must be the case that:

$$-c + \eta\tau_i(h^t, a_c) + \delta E[U(h^t, a_c)|a_i = a_c] \geq \delta \max\{E[U(h^t, a_0)|a_i = a_c], U_0; \} \tag{1.6}$$

where $E[U(h^t, a)|a_i = a_c]$ is the agent's expected continuation surplus having taken the action $a$, given that task $i$ requires the action $a_i = a_c$. The bonus payments promised by the principal must also be credible, so

$$-\tau_i(h^t, a_c) + \delta E\Pi[h^t, a_c|a_i = a_c] \geq \Pi_0. \tag{1.7}$$

In the game with full information, these constraints hold with equality at $\delta = \delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}$, when the continuation value is $U_0$. That implies costly effort can only be obtained in the game with private information if $E[U(h^t, a_0)|a_i = a_c] \leq U_0$. However in the game with private information the principal's ability to set $E[U(h^t, a_0)|a_i = a_c]$ is limited. This is because the principal cannot distinguish between a deviation and the agent correctly (but unsuccessfully) taking the costless action; moreover the principal must still provide incentives on the remaining fraction $\gamma_{-1}(1 - \alpha_i)$ of costly tasks. If he cannot commit to a transfer $\tau \geq c/\eta$, then in order to provide these incentives the principal will need to set the agent's continuation value when task $i$ is in fact costless to be greater than her outside option: $E[U(h^t, a_0)|a_i = a_0] > U_0$. We must have $E[U(h^t, a_0)|a_i = a_0] > E[U(h^t, a_0)|a_i = a_c]$, since by *deviating* to the costless action the agent forgoes any transfers on this task following a success. However for $\alpha$ small these transfers contribute little to the agent's future surplus, and so it will not be possible to set $E[U(h^t, a_0)|a_i = a_c] \leq U_0$ without reneging on the surplus already promised to the agent on other tasks.

This problem can be ameliorated by using the threat of termination. Assume that if the agent takes a costless action and is unsuccessful, then the principal can commit to terminating the relationship with probability $\xi(h^t)$. In this case cooperation requires that:

$$\begin{aligned} -c + \eta\tau_i(h^t, a_c) + \delta E[U(h^t, a_c)|a_i = a_c] &\geq \\ (1 - \xi(h^t))\delta \max\{E[U(h^t, a_0)|a_i = a_c], U_0\} &+ \xi(h^t)\delta U_0. \end{aligned} \tag{1.8}$$

Clearly, by setting $\xi(h^t) = 1$ this incentive constraint reduces to that in the game with full information, ensuring that the agent will take the costly action if appropriate. The trade-off here is that since $\eta < 1$ there is a positive probability that the relationship will be terminated inefficiently on the equilibrium path; moreover this possibility reduces expected surplus earlier in the game, increasing the probability of termination. The resolution of this trade-off will depend on $p$ and $\eta$: for $\eta$ high, the expected loss of surplus due to inefficient termination is small; meanwhile for $p$ large, the possibility of being unable to obtain effort on a costly task carries more weight that the risk of inefficient termination on a costless task.

The above result focused on the case in which it is common knowledge that there is sufficient surplus to sustain cooperation on another costly action on the final tasks, but nonetheless it might not be possible to provide the correct incentives for the agent to undertake this action. We will obtain a similar result, although for different reasons, for $\delta$ close to $\delta^*_{\gamma_{-i}(1-\alpha_i)}$. In this case both parties know that there is insufficient surplus to sustain cooperation on all tasks if the final task requires the costly action. However if the final task is costless then there is sufficient surplus to maintain the relationship with the agent working on all tasks, and receiving a transfer following a success on any of them. Given that the agent obtains a surplus on costless tasks in this case, she will have an incentive to try and deceive the principal as to the true value of the relationship by deviating to the costless action. This implies that if the agent takes the costless action and is unsuccessful then the principal cannot be certain of the surplus from the relationship. At $\delta = \delta^*_{\gamma_{-i}(1-\alpha_i)}$ the expected surplus is such that all incentive constraints bind in the game with full information: introducing any uncertainty will reduce the principal's expected surplus, and hence the bonus that he can commit to pay. This could lead to the breakdown of the relationship even if task $i$ was in fact costless. For larger values of $\delta$, slack in the incentive constraints allows the principal to increase the transfer paid on costly tasks whilst reducing that on costless tasks, removing the incentive for the agent to misrepresent total surplus and hence removing the need for the principal to punish the agent if the costless action is taken unsuccessfully.

**General case in which the value of the relationship is uncertain.**

The previous subsection analysed the problem of providing incentives on a single task assuming that the correct actions on other tasks were known. I now turn to a general case in which there may be multiple tasks that remain unverified. In addition to the problems of incentive provision shown above, there is now uncertainty on both sides about the future value of the relationship. This uncertainty may make it impossible to provide appropriate incentives, even if $C$ is in fact such that full cooperation could be achieved.

To illustrate this, consider the example above, and assume that at some period $t'$ the penultimate task $j$ is drawn, and requires a costly action. As above we can look at different scenarios based on the history of the game and the principal's belief about $\gamma$. If the tasks already learned include a sufficiently high fraction of costless tasks, then the surplus of the relationship will not be an issue, although the principal must still provide the necessary incentives, as above. Alternatively learning that task $j$ is costly may reduce the expected surplus to a level at which full cooperation cannot be achieved regardless of the cost of the final task, and so the agent can never be induced to take the costly action. The final case is the one in which $\delta$ is such that the relationship can be sustained under full information if one of the remaining tasks is costly, but not both. In this case if the probability that task $i$ requires the costly action is sufficiently high, it will not be possible to induce cooperation on task $j$.

In order to ensure that the agent is willing to supply costly effort on task $j$ we require that:

$$-c + \eta\tau_j(a_c, h^t) + E[U(h^t, a_c)|a_j = a_c] \geq \max\{\delta U_0, 0 + E[U(h^t, a_0)|a_j = a_c]\}$$

and moreover that the principal must be willing to pay the transfer, if appropriate

$$-\tau_j(a_c, h^t) + E[\Pi(h^t, a_c)|a_j = a_c] \geq 0.$$

Combining these expressions, given that the agent's continuation value following a deviation must be bounded below by $U_0$, we require that

$$c + \delta U_0 \leq \delta[E[U(h^t, a_c)|a_j = a_c] + \eta E[\Pi(h^t, a_c)|a_i = a_c]] < \delta E[TS|a_j = a_c, h^t]$$

This expected surplus, given that the cost of task $i$ is still unknown, is:

$$E[TS|\gamma_{-i}, \alpha_i] = \frac{1}{1-\alpha_i\delta}\left[(1-\alpha_i)(\eta - \gamma_{-i}c) + \alpha_i(1-p)\left[\eta + \frac{\delta}{1-\delta}(\eta - \gamma_{-i}(1-\alpha_i)c)\right]\right] + \frac{1}{1-\alpha_i\delta}\left[\alpha_i p\left[(1-\mathbb{I}_{quit})\left[\eta - c + \frac{\delta}{1-\delta}(\eta - (\gamma_{-i}(1-\alpha_i) + \alpha_i)c)\right] + \mathbb{I}_{quit}\delta U_0\right]\right]$$

If task $i$ is costless, then if the relationship has survived until this point then it will not be terminated now. If not, the relationship may survive with cooperation on the costly task, or it may be terminated. If $\delta$ and $\gamma_{-i}$ are such that the relationship will survive then this expression simplifies to

$$E[TS|h^t] = \frac{1}{1-\delta}\left[\eta - (\gamma_{-i}(1-\alpha_i) + p\alpha_i)c\right]$$

This is exactly the expected surplus that would be generated in the game with full information if a fraction $\gamma_{-i}(1-\alpha_i) + p\alpha_i$ of tasks were known to be costly. However if $\delta$ is such that the relationship cannot be sustained if both tasks $i$ and $j$ are costly, then the relationship will terminate if the agent learns that the last task is costly. This implies that the expected surplus in the game with private information is strictly less than that in the game with full information, for any $p > 0$:

$$E[TS|\gamma_{-i}, \alpha_i] < E[TS_{full\ info}|\gamma_{-i}, \alpha_i] \leq \frac{1}{1-\delta}\left[\eta - (\gamma_{-i}(1-\alpha_i) + p\alpha_i)c\right].$$

In particular, this means that it will be impossible to induce the agent to supply effort for $\delta = \delta^*_{(1-\alpha_i)\gamma_{-i}}$ if there is any chance that the last task $i$ will also be costly. It follows that there are cases in which this task is in fact costless, and so cooperation could have been supported under full information, but in the game with private information the relationship will be terminated. Moreover, in such a case, had task $i$ been required before task $j$, this would have ensured enough surplus to maintain the relationship. Thus the order in which tasks are required will be critical to the long-run performance of the partnership.

We can apply this same argument for each successive task as it is first drawn. This yields the following proposition.

**Proposition 4.** *In the game with full information, let $\delta^*_p$ be the discount factor required to sustain full cooperation when a fraction p of tasks require the costly action. In the game with private information in which the independent probability that any particular task is costly is p, a discount factor $\hat{\delta}(p) > \delta^*_p$ for all $p > 0$ is required for cooperation if a costly task is drawn at the start of the game.*

This result states that there exist values of $\delta$ for which it is possible to achieve full cooperation in the game with full information when a fraction $p$ of tasks are known to be costly, but it may not be possible to achieve this in the game with private information when in expectation a fraction $p$ will be costly. This implies that for some realisations of $C$ such that cooperation is possible in the game with full information, cooperation will only be possible in the game with private information if the tasks are encountered in a specific order.

The proof is in the appendix, but this result can be shown by inductively constructing an upper bound for the expected surplus in the game with private information, following any history. This upper bound coincides with the expected surplus in the game with full information. It can then be shown that for any $\delta < \delta^*_1$ and any $p > 0$ there will exist histories such that the expected surplus given this history is strictly less than this upper bound, implying that cooperation will be possible for a smaller set of parameter values than in the game with full information.

As a corollary to this result, we can see that for any $p > 0$ the relationship runs the risk of inefficient termination on the equilibrium path. Moreover, for $p$ close to one it will not be possible to achieve any cooperation on costly tasks, or to get the agent to participate in the relationship, even for $\delta$ arbitrarily close to $\delta^*_1$. We can see this by noting that at $p = 1$, the agent cannot be induced to take the costly action for any $\delta < \delta^*_1$, and moreover she prefers not to enter the relationship at all, rather than entering for one period and postponing her outside option. Since the expression for surplus is continuous in $p$, this implies that there exist values of $p < 1$ such that the agent will never enter the relationship, even if $\delta$ is very close to the level that would support cooperation even if all tasks are costly.

### 1.4.3 Equilibria with "probation"

Thus far in the paper, we have focused on equilibria with the following characteristic: the principal will terminate the relationship if the agent takes an action $a(i,s) \neq a(i,t)$ at any time $s > t$. That is, the agent cannot switch from one action to another on a given task without the principal terminating the relationship. In this section of the paper I relax the assumption that the agent is expected to provide appropriate effort on all tasks from the outset. I show that in this case, since the agent is no longer required to take costly actions until she can acquire more information, the set of discount factors for which cooperation can be achieved is increased.

We could think of such an equilibrium as allowing for a probationary period. The principal would permit the agent to take only costless actions while she ascertains the true value of the relationship, after which, if there is enough surplus, the agent can start to work on costly tasks. Whilst the agent may still be promised a bonus if she takes a costly action at the start of the game, she will not be punished if she takes the costless action on this task initially and later switches to the costly action. In order for such a policy to succeed in inducing cooperation on all tasks (wherever possible in the game with full information), two conditions must be met. Firstly, the expected surplus during the probationary period must be sufficient to ensure both parties are willing to participate and forgo their outside options. Secondly, the agent must be induced to take the costly action, when appropriate, after the probationary period is over. I show that the set of discount factors for which the agent will remain in the relationship will increase, reducing the probability of inefficient termination due to uncertainty about the surplus. However an equilibrium of this type cannot resolve the incentive issues shown in Proposition 3, and the agent may still quit the relationship inefficiently after certain histories.

**Proposition 5.** *For any $p > 0$, assume that $\delta$ is such that the agent can be induced to take a costly action in the equilibrium described in section 4.2. Then cooperation can also be ensured in the equilibrium with probation. Moreover the set of values of $\delta$ for which cooperation can be achieved on a costly action at the start of the game will strictly increase.*

*Proof.* Since the equilibrium with probation does not rule out a case in which the principal promises the same transfers as in the baseline equilibrium, and the agent immediately takes the costly action whenever these transfers elicit effort in that equilibrium, it is clear that the equilibrium with probation can replicate any payoffs achievable in the equilibrium without probation. I now show that there will exist histories following which the equilibrium with probation does strictly better. Given that these histories will occur with positive probability, allowing a probationary period must increase the expected surplus at the start of the game, thereby increasing the set of values of $\delta$ for which the agent can be induced to take a costly action at the start of the game.

As shown above, in the equilibrium in which the agent is expected to fully cooperate from the start of the game, she will only take the costly action on a given task if

$$-c + \eta\tau_i(a_c, h^t) + \delta E[U(h^t, a_c)|a_i = a_c] \geq U_0.$$

If we relax this constraint on the agent's behaviour, we require that the agent is willing to either take the costly action, or to take the costless action, and remain in the relationship until she has acquired more information. The relevant constraint becomes:

$$\max\{-c + \eta\tau_i(a_c, h^t) + \delta E[U(a_c, h^t)|a_i = a_c], 0 + \delta E[U(a_0, h^t)|a_i = a_c]\} \geq U_0.$$

Given that the agent always has the option of playing as if in the baseline equilibrium in all subsequent periods and receiving the concomitant payoffs, we must have $E[U(a_0)|h^t, a_i = a_c] \geq E[U(a_c)|h^t, a_i = a_c]$. Given that a transfer $\tau = c/\eta$ is sufficient to achieve cooperation on any task, without requiring the principal to commit future surplus, the principal will never promise a transfer $\tau > c\eta$. Moreover following certain histories we must have $\tau < c/\eta$. This follows from the fact that the discount factor required to sustain cooperation is minimised by cross-subsidizing tasks, which must mean that $\tau < c/\eta$ for any $\delta < \delta_1^*$. Given that $C$ is initially unknown, for any $\delta$ there is a positive probability of reaching a history in which cooperation can only be achieved by setting $\tau < c/\eta$. This implies that we must have

$$0 + \delta E[U(a_0)|h^t], a_i = a_c] > -c + \eta\tau_i(h^t, a_c)$$

following certain histories, and therefore that the set of values of $\delta$ for which the agent can be induced to remain in the relationship strictly increases. By reducing the probability that the relationship is terminated abruptly when there is in fact sufficient surplus to continue the relationship, the expected surplus of the relationship increases, making it easier to achieve cooperation at the start of the game.

□

Therefore relaxing the restriction that the agent cannot switch actions on a given task can increase the set of values of $(\delta, p)$ for which the agent can eventually be induced to take the costly action. However this will not necessarily be sufficient to replicate the payoffs in the game with full-information for any $U_0$. This is partly because this equilibrium cannot resolve the problem of providing correct incentives since the extent to which the principal can manipulate the continuation surplus remains constrained. However as long as the agent has a positive outside option, then for sufficiently large $p$ there will still be histories after which the agent quits rather than remaining in the relationship to acquire more information.

To see this, consider a case in which cooperation can be ensured on all tasks if a fraction

$1-\gamma$ are costless, that is $\delta \geq \delta^*_\gamma$. In the game with private information, if the agent has observed that a fraction $1-\gamma$ of tasks are costless, then she will be willing to take the costly action on other tasks, as required, since there will be no risk that the surplus from the relationship drops below the level necessary to ensure cooperation in the future. However, until a fraction $1-\gamma$ of tasks are known to be costless, the agent will not take the costly action, but must accept a payoff of zero on those tasks until she obtains further information. We therefore require that there is sufficient surplus to ensure that the agent stays in the game until more information is acquired.

Consider the following potential history. Whilst a fraction $1-\gamma$ of tasks are in fact costless, the agent encounters the fraction $\gamma$ of costly tasks before encountering any costless tasks. In this case the likelihood that the relationship will terminate increases, and the expected continuation surplus from the game decreases. Moreover, since the agent cannot be induced to take the costly action (which does generate an expected surplus) until she has more information, the potential gains from the relationship are deferred and the agent receives $0 < U_0$ on the costly tasks in the interim. In this case the expected surplus from remaining in the probationary stage until she receives more information will drop below $U_0$ for sufficiently high $p$, and therefore the agent will terminate the relationship.

Given that the history described above will occur on the equilibrium path with positive probability for certain $(\delta, p)$, in the game with private information the relationship may break down on the equilibrium path for values of $\delta$ for which cooperation would be possible in the game with full information. Relaxing the restrictions on the principal's equilibrium strategy therefore cannot generate the first-best for any $U_0 > 0$.

## 1.5 Example

In this section I provide some illustration of the results of the previous section, in particular those relating to the dispersion in firm performance and the dependence of long-run payoffs on early draws. For purposes of exposition, I will do this through two distinct examples. The first example involves just two tasks, which occur with equal probability. In this example the only source of relationship breakdown is the *ex ante* uncertainty about whether the tasks will be costly, and hence about the future value of the relationship. Following proposition 4, I show that this uncertainty may make full cooperation impossible in cases where it could be achieved under full information. I then consider the equilibrium with probation, and use this setting to explicitly show the extent to which allowing probation improves the expected long-run value of the relationship, as discussed in section 4.3. This example also illustrates that long-run performance may depend critically on the order in which tasks are drawn: if one task is costless and the other costly, there are values of $\delta$, $p$ for which drawing the costless task first guarantees success, but drawing the costly task first (given that neither party knows the cost of the other task) will lead to the breakdown of the relationship.

The second example has three tasks, which occur with asymmetric probabilities. In this case, in addition to the uncertainty about the surplus generated, there will also be problems of incentive provision following certain histories. Following Proposition 3, I show that there are regions of the parameter space in which, even if both parties know that there is sufficient surplus to achieve full cooperation, the principal cannot provide the appropriate incentives without using the threat of (inefficient) termination. Moreover, this example shows that it will be more difficult to provide incentives on tasks that occur infrequently. The intuition for this is that the upfront cost $c$ faced by the agent on such tasks is the same, but the future value of effort, and hence the surplus foregone by deviating, is smaller for tasks that are rarely required. As in the two-task example, the long-run payoffs will depend not only on the realised value of $C$ but also on the order in which tasks are first encountered. I conclude this section with an illustration of the heterogeneity in outcomes for firms facing identical draws of costs across tasks, but encountering the tasks in a different sequence, and with differening success.

### 1.5.1 Example with two tasks

This example focuses on the results in propositions 4 and 5. That is, I show that full cooperation may not be possible in cases in which it is possible in the game with full information. I go on to consider the equilibrium with probation or cooperation on only a subset of tasks, to illustrate the extent to which this can improve the expected long-run performance of the partnership. We can see this by comparing the areas of the parameter space $(\delta, p)$ for which full cooperation is possible in each equilibrium, which will be shown in figures 1 and 2. In both cases the long-run outcome can depend on the order in which tasks are required. The details of this example are in the appendix.

Assume that the principal and the agent are required to work on two tasks, which occur with equal probability. Let the principal's outside option $\Pi_0 = 0$, and let the agent's outside option be $U_0$. From the full-information case, we may obtain the set of critical values of $\delta$, $\{\delta_i^*\}_{i=0,1/2,1}$ for which full cooperation is possible, and the cutoff $\delta_{1/2}^0$, at which the agent can be persuaded to remain in the relationship, working only on a single, costless, task. Note that $\delta_0^* < \delta_{1/2}^0 < \delta_{1/2}^* < \delta_1^*$.

*Cooperation from the start of the game*

I first consider the baseline equilibrium in which the agent is expected to cooperate, and take costly actions if necessary, from the start of the game. If the first task to be drawn requires the costless action, then there is no difficulty in providing incentives. The principal must only ensure that the agent is willing to remain in the relationship, and if the expected surplus was sufficient to induce the agent to participate at the start of the game, it must be sufficient to induce the agent to remain in the game given that the expected value of the relationship has increased.

Moreover if the first task is costless then the principal can obtain costly effort on the second task, if necessary, whenever he would be able to do so in the case with full information. The details are in the appendix, but the argument is that the principal can hold the agent to her outside option if both tasks are in fact costless, meaning that if she *deviates* to the costless task she must receive less than her outside option, since she will never receive any transfers. The principal can then offer the same transfers as in the full information case if she takes the costly action: we know that these transfers are incentive compatible for the principal, and are preferred by the agent to her outside option. Therefore cooperation is assured.

It is worth noting that for $\delta$ close to $\delta^*_{1/2}$ the transfer paid on the costless task will increase after the costly task has been confirmed: the principal wants to minimise the surplus given to the agent if both tasks are costless in order to use the promise of greater surplus to provide incentives should the second task be costly. He can do this be paying larger transfers to the agent on both tasks once the costly action has been verified on the second task.

If the first task required is costly, however, then maintaining the relationship becomes more difficult. If $\delta < \delta^*_{1/2}$ the agent can never be persuaded to take a costly action. Otherwise, whether or not it will be possible to persuade the agent to cooperate and take this action will depend on the probability that the second task will also be costly. Assume that $\delta \in [\delta^*_{1/2}, \delta^*_1)$, then, as shown in Proposition 4, it will no longer be possible to induce the agent to take the costly action at $\delta = \delta^*_{1/2}$ for any $p > 0$. Moreover, as $p$ increases it will reach a point at which cooperation cannot be achieved even if $\delta$ is arbitrarily close to $\delta^*_1$. The range of parameter values for which cooperation can be achieved is shown in figure (2).

Given the difficulty of inducing the agent to cooperate on a costly task at the start of the game, we can see that the eventual surplus generated by the relationship may depend crucially on the order in which tasks are first required. In particular, at $\delta = \delta^*_{1/2}$, the principal cannot induce the agent to take the costly action at the start of the game for any $p > 0$, however if the costless action is required first then we are assured of obtaining full cooperation. In this case, the breakdown in the relationship arises entirely from the uncertainty about the total value of the relationship. Moreover, this breakdown is due to the fact that the principal is trying to induce the agent to take a costly action before this uncertainty is resolved. In light of this, it would seem sensible for both parties to agree that the agent should not have to undertake any costly tasks until she knows more about the value of the relationship. I now consider the extent to which equilibria with probation and equilibria which allow for partial cooperation on only one (costless) task can do better.

*Equilibrium with probation.*

In order to illustrate the extent to which allowing probation can improve the likelihood that the relationship survives, I relax the assumption that the agent must take the costly action, if required, from the beginning of the game. Instead assume that the agent may be permitted to take the costless action on any task until she has established that there is sufficient surplus to support the transfers necessary to induce cooperation on a costly task, or, failing that, she may continue to work only on costless tasks throughout. I show that for some parameter values allowing probation can ensure that the relationship reaches the full-information level of output, when the agent would otherwise have quit. In this case we may see substantial short-run heterogeneity in firm output, as the duration of the probationary period varies. Moreover there will still be regions of the parameter space in which the draw of the first task is critical for the survival of the partnership.

There are two issues we must consider when evaluating the feasibility of the equilibrium with probation: can the agent be persuaded to remain in the relationship, forgoing her outside option, while she collects information, and can the agent be induced to take the correct actions once she is in possession of this information. In the two-task case the second consideration will not prove a problem: if only one task requires the costly action then the agent can be induced to take this action for any $\delta \geq \delta^*_{1/2}$. This can be achieved with the same transfers as in the baseline equilibrium.

We now need to ascertain whether the agent will be willing to remain in the relationship to acquire information. Specifically, if the first task required is costly, then even though the agent is not required to take the costly action her belief about the value of the relationship may nonetheless drop to a level at which she prefers to quit. We can consider different cases, for different values of $\delta$. If $\delta \in [\delta^*_0, \delta^0_{1/2})$ then the agent will only remain in the relationship if both tasks are costless, and therefore will quit as soon as a costly task is drawn; moreover for higher values of $p$ she will not be willing to enter the partnership at all. For $\delta > \delta^*_1$ then we know that there is sufficient surplus to support cooperation on any combination of tasks, from the outset, and so there is no need for probation. We can now consider the two intermediate cases.

For $\delta \in [\delta^0_{1/2}, \delta^*_{1/2})$ the agent cannot be persuaded to work on a costly task, but there is sufficient surplus to support the agent working on a single costless task. Therefore if the costless task is drawn first the agent will remain in the relationship, regardless of the cost of the second task. If, however, the costly task is drawn first her expected surplus from the relationship will fall. In a result similar to that of proposition 3, we find that at $\delta = \delta^0_{1/2}$ the agent will not be willing to stay in the relationship, for any $p > 0$. The intuition is that for this value of $\delta$ the agent is indifferent between staying in the relationship and taking her outside option in the full information case, given the largest transfer the principal can credibly pay. Any uncertainty about future surplus means that the agent will prefer to quit. Nonetheless, compared to the equilibrium without probation, in which for $\delta < \delta^*_{1/2}$ the agent would immediately quit after drawing a costly task, the set of values of $\delta$, $p$ for which the relationship may continue strictly increases.

The other intermediate case to consider is $\delta \in [\delta^*_{1/2}, \delta^*_1)$, in which it is possible to support full cooperation if just one task is costly. If the costless task is drawn first then, as before, the relationship is secure; otherwise, the agent will only remain in the relationship if her expected payoff, conditional on first drawing a costly task, is sufficiently high. The expected surplus from remaining in the relationship is the same as in the example without probation: once the agent learns the second task, she will either quit if it is costly, or work on both tasks if it is costless. The difference comes from the fact that in the interim the agent is not expected to provide costly effort, and so the agent is only concerned with the opportunity cost of forgoing her outside option. Therefore this strictly increases the set of parameter values for which the agent is willing to remain in the relationship after drawing a costly task at the start of the game.

*Parameter values for which cooperation can be achieved.*

We can put these results together to compare the parameter values for which it is possible to induce the agent to cooperate, depending on whether or not the first task drawn is costly or costless.

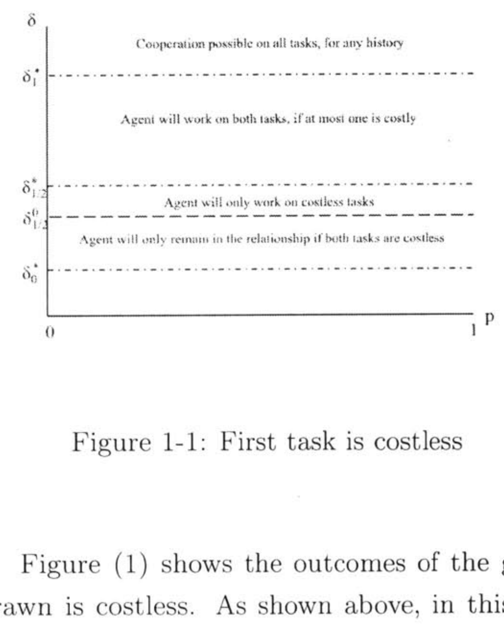


Figure 1-1: First task is costless

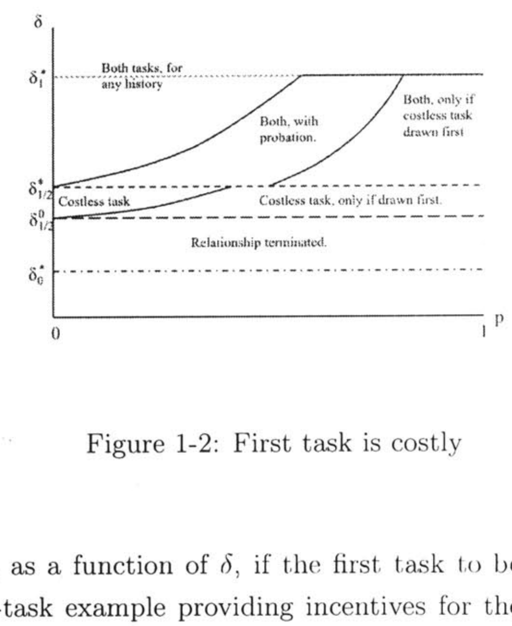


Figure 1-2: First task is costly

Figure (1) shows the outcomes of the game, as a function of $\delta$, if the first task to be drawn is costless. As shown above, in this two-task example providing incentives for the agent on the second task will not be a problem if the first task is costless. Therefore the principal will be able to obtain the correct action from the agent whenever it would be possible to do so in the game with full information.

Turning to figure (2), we see that if the first task to be drawn is costly then obtaining cooperation becomes more difficult. In particular, if the principal insists that the agent cooperate from the outset, the relationship will be maintained only for small values of $p$, and cannot survive at $\delta = \delta^*_{1/2}$ for any $p > 0$. Relaxing the requirements on the equilibrium to allow the agent avoid supplying costly effort until she has acquired more information increases the set of parameter values for which full cooperation is possible, but nonetheless there will still be regions of the parameter space in which full cooperation can be achieved if and only if the costless task is drawn first.

From this simple example we can already see that there will be heterogeneity in firm performance, even conditioning on the realised draws of tasks. This will occur both in the short-run, as the time it takes for firms to move from probation to full cooperation varies, and in the long-run, if some partnerships terminate altogether.

### 1.5.2 Example with three tasks

In the example with just two tasks the breakdown in the relationship is due entirely to the uncertainty about the long-run value of the relationship: once the agent knows that there is sufficient surplus to sustain cooperation the principal can provide the incentives necessary to achieve this. In this example I focus on the problem of providing the necessary incentives, as per Proposition 3. The critical difference from the two-task example is that once more tasks are introduced we may see a situation in which the principal wishes to ensure cooperation on a particular task, having already learned that a prior task is costly. Inducing the agent to take the costly action on an earlier task requires guaranteeing the agent a certain share of future surplus, reducing the principal's ability to use this surplus to induce cooperation on other tasks. This will generate an additional source of heterogeneity in long-run performance: in the two-task case, the parties would either achieve the optimal full-information level of performance, or terminate the relationship; in this case we may also see firms achieving an intermediate level of output by working on a subset of tasks. This example also demonstrates how the difficulty in obtaining effort will vary with the frequency with which tasks occur.

Assume now that the principal and the agent work on three tasks, $\{A, B, C\}$, in which tasks are required with differing frequency. Let the weights on each task be $\{\alpha_i\}_{i=A,B,C} = \{0.45, 0.45, 0.1\}$. Given the possible draw of costs, we will obtain thresholds $\{\delta^*_{0.1}, \delta^*_{0.45}, \delta^*_{0.55}, \delta^*_{0.9}, \delta^*_{1}\}$. I will begin by considering the continuation game after two tasks have been verified, and show that it may not be possible to provide the incentives necessary to elicit costly effort. I then turn to cooperation in earlier stages of the game, and show that uncertainty about the true value of the relationship may prevent the principal from offering adequate incentives for effort, analogously to the two-task example. As in that case, allowing for a probationary period ameliorates the effects of uncertainty, but does not resolve the issue of incentive provision. The last part of this section shows that the performance of the partnership may vary substantially even for a fixed draw of $C$, due to the order in which tasks are required.

*Cooperation on the third task*

Assume that the actions required on two of the three tasks have already been verified. If either both actions are costly, or both costless, then we can be assured that cooperation will be possible on the third task whenever it is possible in the game with full information.

This situation is analogous to that in the two-task example, but with asymmetric weighting on the tasks. However we may see a different outcome if one of the verified tasks is costly, and the other costless. In this case the fact that some surplus must be guaranteed to the agent in order to ensure cooperation on the prior costly task limits the principal's ability to use the promise of surplus to induce effort on other tasks.

We can consider two cases here: that in which the third task carries weight 0.1, and that in which it carries weight 0.45. Assume in both cases that Task B, which occurs in any period with probability 0.45, is costless. Let $\delta = \delta^*_{0.55}$, and so it is common knowledge that there is sufficient surplus to sustain cooperation in the game with full information. Assume that the second task required, either Task A or Task C, is known to be costly. The principal needs to provide incentives that will induce the agent to take the costly action on the third task, if necessary. Following the argument of Proposition 3, the principal can induce the agent to supply costly effort on the third task if the agent's continuation value following a deviation can be set less than or equal to her outside option. However the principal cannot distinguish between a deviation on the third task, and the agent correctly taking the costless action but without success. In the latter case, the principal wants to ensure that the agent will continue to work on the other costly task. The need to provide incentives on previously attempted tasks and the inability to distinguish whether the agent has deviated therefore constrain the principal's ability to manipulate the agent's continuation value.

This problem of incentive provision will not occur on tasks that are required frequently, in this case, on Task A, where $\alpha_A = 0.45$. Intuitively, this is because the principal can set transfers such that the difference between correctly taking the costless action on Task A (but without success), and *deviating* to the costless action is large. Given that the principal cannot distinguish between these two cases, the difference in the continuation values to the agent comes from foregone transfers: the agent knows that if she deviates she will not receive any transfers on Task A in the future. Given that Task A occurs frequently, this represents a substantial drop in the agent's continuation surplus. This means that the continuation value following a deviation can be set to be less than the outside option, whilst still ensuring the agent has enough surplus to cooperate if the third task is in fact costly.

If the third task occurs less frequently this will not be possible. Assume that Task A is already verified as costly, but the principal must provide incentives on Task C. Since Task C occurs infrequently, transfers paid on this task contribute little to the agent's total surplus. Therefore any attempt to set the continuation value following a deviation to be less than the agent's outside option must also greatly reduce the agent's continuation value when she correctly takes the costless action but is unsuccessful. Moreover, since Task A is costly and occurs frequently, more surplus is required to ensure cooperation on this task in the future. In this case the principal must either accept that the agent will not be willing to provide costly effort on Task C, or he can use the threat of termination following a failure on this task to provide incentives. For $\eta$ large, the probability of the agent correctly taking this action but being unsuccessful is small, and so this may be optimal.

*Cooperation on earlier tasks*

We can now turn to cooperation in earlier stages of the game, taking into account the fact that after certain histories it may not be possible to obtain costly effort on later tasks, even if the discount factor is sufficiently high to ensure full cooperation in the game with full information. The results here are similar to those in the example with two tasks (for more detail, see the appendix), but it is worth noting how the weight attached to different tasks affects the likelihood that the relationship breaks down.

If we consider the subgame after one task has been verified, then we can find a lower bound for the discount factor required to sustain cooperation on a specific costly task. In this case it will also depend on the task already verified and on the weight attached to the new task. For example, assuming that Task B is verified as costless, we could compare the set of discount factors for which it is possible to persuade the agent to take a costly action on Task A and Task C respectively. Given that Task A is required more often than Task C, learning that Task A is costly leads to a larger drop in the agent's expected surplus, making it more difficult to induce the agent to supply costly effort. However these tasks differ not only in the probability they are required, but in the outcomes of the continuation games: even if the agent can be persuaded to supply costly effort on Task A, we know that it might not be possible to induce her to cooperate on Task C. If Task C is verified as costly then we know that cooperation can later be sustained on Task A whenever it is possible to do so in the game with full information.

The same argument can be made with regard to cooperation on the first task drawn: in the early stages of the game it is easier to induce the agent to supply effort on tasks that carry little weight. This is due partly to the direct effect that learning such a task is costly has a smaller impact on expected surplus. There is also an indirect effect due to the fact that in the later stages of the game it is harder to induce costly actions on tasks that occur infrequently; encountering these tasks early on rules out the possibility of this type of inefficiency arising later. We can also see that if the principal and the agent could choose the order in which tasks were attempted (before learning their costs), they would face a trade-off: learning that a frequently occuring task is costless early on may ensure that there is enough surplus to sustain the relationship, but this is risky if the task turns out to be costly. Attempting infrequent tasks early on reduces the risk of the relationship ending if the task is costly, but learning that the task is costless may not have much impact on the continuance of the relationship.

As in the two-task example, allowing probation will increase the set of parameter values for which it is possible to eventually reach full cooperation, however it does not eliminate the possibility that the relationship will break down following certain histories. Moreover, in this case, it cannot solve the problems of incentive provision and therefore we would still expect to see inefficiently low performance after certain histories.

### 1.5.3 Path Dependence

Following the analysis above we can consider the variety of outcomes which may occur on the equilibrium path for different draws of costs $C$. The figures in this section show the dispersion in possible outcomes conditional on a fixed draw of $C$, illustrating that even if firms face identical tasks and costs there will still be variation in their performance, both in the long-run and as they learn $C$. I focus on equilibria that allow a probationary period, i.e., a period of only partial compliance. The differences in performance become more stark if we rule out this option. If we consider the baseline equilibrium in which the principal seeks to enforce full cooperation from the start of the game then it is not possible for the relationship to recover from a period of partial cooperation: it will either remain at this level, or the relationship will be terminated.

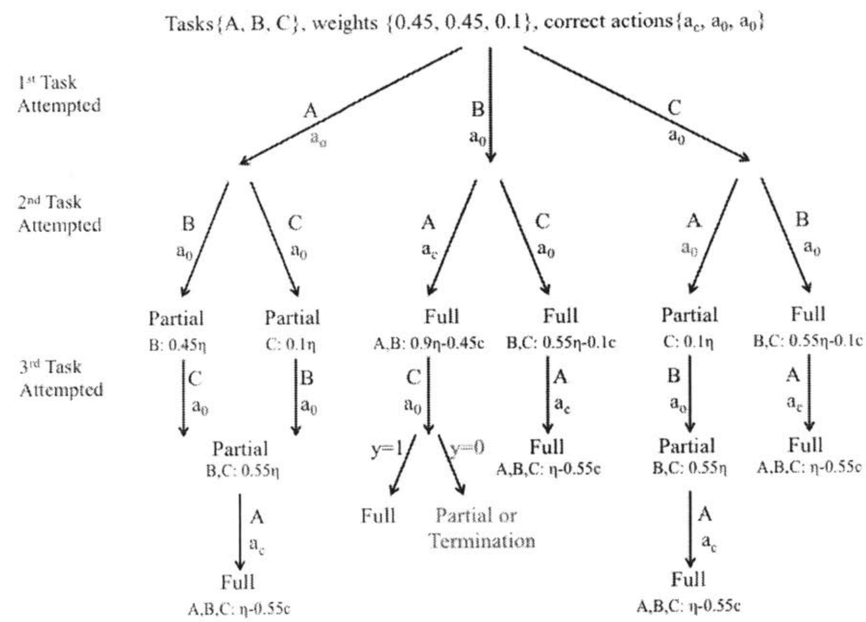


Figure 1-3: Path Dependence when two tasks are costly

Figure (3) illustrates the dispersion of outcomes on the equilibrium path when the actualised costs are $\{a_c, a_0, a_c\}$, that is, when a fraction $\gamma = 0.55$ of tasks require the costly action. The figure shows the various orders in which the tasks may first be required, and the action that the agent will take on each task - which may not be the correct action, especially when the task is first required. The figure also shows the expected payoffs from working on a limited set of tasks, for example during the probationary period. Note that until all tasks have been drawn neither the principal nor the agent know $C$, thus their strategies must take into account all possible cost combinations, even though only the histories that can follow a particular draw are shown here. Assume that $\delta = \delta_0^*.55 + \varepsilon$, and $p = 0.5$: this implies that it is possible in the game with full information to achieve cooperation on all tasks. We can see that the eventual long-run level of performance will depend on the first task drawn. For example, drawing Task A at the start of the game is damaging to the relationship: not only is it impossible to obtain costly effort on this task from the agent, but if Task C is then drawn before the costless Task B the agent will become so skeptical about the value of the relationship that she will quit altogether.

In contrast, drawing Task C first will ensure that the partnership will eventually be able to work at the optimal level: for these parameter values, even if Task A is drawn before Task B, the fact that some surplus can be generated on Task C in the interim will convince the agent to remain in the partnership. Moreover, drawing Task C at the start of the game removes the risk of being unable to ensure cooperation on this task later in the game.

We can compare these outcomes with those when the draw of costs is $\{a_c, a_0, a_0\}$, i.e. when only $\gamma = 0.45$ of tasks are costly, for the same values of $\delta$ and $p$. This is shown in figure (4). For these parameter values the risk of the agent quitting the relationship is removed: whilst she will not be willing to take the costly action on Task A at the start of the game, she is willing to remain in the relationship until it is clear that there is sufficient surplus to proceed. However we may still see suboptimal performance due to the fact that the principal cannot be sure of the true value of the relationship on observing the agent take a costless action on Task C without success. This generates uncertainty about the true value of the relationship, which undermines the credibility of the promised transfers. This may lead to partial cooperation on only the costless tasks (until a success is achieved on Task C, and the principal can verify the surplus from the relationship), or to termination if the principal reneges on a transfer.

Whilst it is for these intermediate values of $\gamma$ that we will see the greatest heterogeneity in performance, we can also consider the other possible draws of costs here. If all tasks are costless, then cooperation is assured, and the order in which tasks are required is inconsequential. Likewise if only Task C (with weight 0.1) is costly: for the parameter values considered here it will be possible to induce effort on this task following any history. If all tasks turn out to be costly, then in the long-run no cooperation is possible; however if Task C is drawn first then the agent may be persuaded to supply effort on this task for a limited period before eventually terminating the relationship; if either Task A or Task B is drawn first the agent will not supply any effort, and the relationship is likely to terminate sooner.

In general, therefore, we can see that not only are there short-run differences in performance as partnerships learn the value of future surplus and the correct actions on each task, but we may still see long-run differences in performance if it is impossible to provide appropriate incentives after certain histories. Considering the evolution of this heterogeneity over time, we will expect the variance in output to decrease, but not disappear, even for firms facing identical costs $C$.

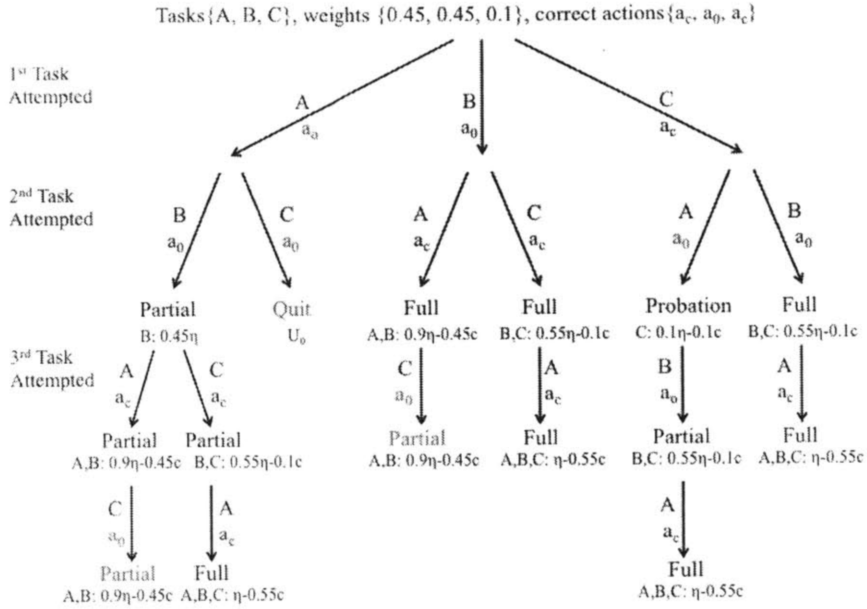


Figure 1-4: Path Dependence when one task is costly

## 1.6 Persistent Performance Differences

Thus far, this paper has shown that success in maintaining the relationship in this model, and the level of productivity achieved by the partnership may be highly dependent on the order in which tasks are first required and on whether early attempts on a particular task are successful. We may therefore expect to see great variation in outcomes across partnerships, even if the underlying distribution of costly tasks $C$ is the same. This may explain why seemingly similar firms have very different levels of productivity, and if we consider each partnership in isolation then in the absence of any further information these differences will persist. However, if we consider a larger set of firms, which are known to share the same set of costs $C$, then any individual partnership should be able to infer something about the value of their relationship by observing the performance of other firms. In this case a partnership that has initially performed poorly may be able to recover if it becomes clear from others' performance that there is in fact enough surplus to maintain the relationship. Over time, therefore, these differences in performance might be expected to disappear.

In this section of the paper I ask whether increased knowledge about the total value of the relationship can alleviate the informational problems that prevent full cooperation on all tasks. For simplicity, I will focus on a simple case in which there is an "incumbent" firm which has full information about the cost of each task, and a new firm that has no information. I will assume that the individuals in each firm share a common discount factor $\delta$, and that $C$, $c$ and $\eta$ are known to be equal across firms. There are multiple information structures that could be envisaged in this setting: firstly, the new firm may only be able to observe the fact that the incumbent firm is still in existence, but not their level of output; secondly, the new firm may be able to observe revenues, but not costs; and lastly, that the new firm can observe the profits of the incumbent. I will consider each of these settings in turn.

In the first case, the new firm can observe only that the incumbent is still in production, but does not know whether the agent is working on all tasks, or only on the costless tasks. In some cases this may be informative. For example, for any given $\delta$, learning that a firm remains in the market places an upper bound on the number of costly tasks. Specifically, the new firm can construct the set $\{\delta^0_\gamma\}_{\gamma\in\Gamma}$, where $\Gamma$ is the possible fraction of tasks that may be costly, given the number of tasks and their weights.[1] Knowing that the incumbent shares the same discount factor, and that the relationship has survived, the new firm can infer an upper bound for $\gamma$: $\bar{\gamma} = \max\{\gamma \in \Gamma | \delta^0_\gamma < \delta\}$. This is turn places a lower bound on the potential surplus from the relationship, and knowing this, the agent can be induced to remain in the relationship even if the draws of early tasks are discouraging, at least in the equilibrium with probation. However knowing this information does not alleviate the incentive problems, and we may still see cases in which firms remain in the probationary stage for a long time.

The second case to consider is that in which the new firm can observe the revenues or the level of output of the incumbent, and can thereby ascertain whether that firm is able to work on all tasks, or only a subset of costless tasks. I will abstract away from the fact that learning per-period output over a finite history cannot convey this information with certainty, given that the correct action only generates a success with probability $\eta$. We may assume that the new principal and agent observe a sufficiently long history and can apply the law of large numbers.

As discussed before, either the agent should work on all tasks, or only on the costless tasks; we will not see an equilibrium in which the agent works on a subset of the costly tasks. In the former case, the new firm will observe mean per-period revenues of $\eta$. Once again, given that firms share a common discount factor, this allows the firm to infer an upper bound on the fraction of tasks that require the costly action. Moreover this bound will be more informative than in the previous case. Since the incumbent is working on all tasks, we will have $\bar{\gamma} = \max\{\gamma \in \Gamma | \delta^*_\gamma < \delta\}$, where $\delta^*_\gamma > \delta^0_\gamma$, for all $\gamma \in (0, 1)$. Since this guarantees that the surplus from the relationship is sufficient to sustain full cooperation, the agent will be willing to supply costly effort at the start of the game. Nonetheless, the problems of providing adequate incentives on particular tasks will remain, and so we will still see inefficiently low performance following certain histories.

In the case in which the incumbent is working only on the costless tasks, the new firm can observe $\gamma$ perfectly, since the average per-period surplus will be $\gamma\eta$. As in the first case,

[1] In the three example case above, for instance, $\Gamma = \{0, 0.1, 0.45, 0.55, 0.9, 1\}$.

this will ensure that in an equilibrium with partial cooperation the agent does not quit the relationship if she draws costly tasks at the start of the game. No incentives need to be provided in this case, since the agent will only take costless effort, but the principal must still ensure that the agent chooses to participate. Knowing $\gamma$ with certainty allows the principal to extract all the surplus from the relationship and hold the agent to her outside option. In the case above, in which there is a positive probability that $\gamma < \bar{\gamma}$, the agent would have obtained some of this surplus.

Increasing the amount of information available to the new firm further, we can consider the situation in which the new firm may observe both the revenues and the profits of the incumbent. In this case, subject to the same assumptions that the new firm is able to observe sufficient history to make a reliable inference, the new firm can learn the precise fraction of tasks that will require the costly action, although not the identities of these tasks. This could potentially help the principal enforce the correct actions since he can terminate the game if the agent takes the costless action on too many tasks; a situation such as that described in Proposition 3 can therefore be avoided without the threat of termination on the equilibrium path. However this alone will not resolve all the difficulties: the agent may still be tempted to deviate to the costless action in order to increase her surplus in the short run, even though she knows that this will eventually lead to the termination of the relationship.

To see this, let a fraction $\alpha$ of tasks be commonly known to be costly, and assume that $\delta \geq \delta^*_\gamma$. For $\delta$ sufficiently large, it will be possible to ensure cooperation on all tasks by paying a transfer $\tau \geq c/\eta$ to the agent only following a success on a costly task.This effectively means that the agent is rewarded on a task-by-task basis, and so she cannot gain by deviating to a costless action when this action cannot possibly be successful. For this to be possible, the promised transfer must be credible, that is, we need:

$$-\frac{c}{\eta} + \frac{\delta}{1-\delta}\left[\alpha\eta\left(1-\frac{c}{\eta}\right) + (1-\alpha)\eta\right] \geq \Pi_0.$$

This will hold with equality for some $\delta$ strictly greater than $\delta^*_\gamma$. For intermediate values of $\delta$ close to $\delta^*_\gamma$, in order to ensure that the agent cooperates on all tasks the principal must increase her continuation surplus further by paying a transfer following success on the costless tasks as well. For the agent to be induced to take the costly action on a task $i$, it must be that

$$-c + \eta\tau_i(h^t) + \delta E[U(a_c, h^t)|a_i = a_c] \geq 0 + \delta E[U(a_0, h^t)|a_i = a_c].$$

At $\delta = \delta^*_\gamma$, for any set of transfer payments that are incentive compatible for the principal this expression will bind with equality for $E[U(a_0, h^t)|a_i = a_c] = U_0$, i.e., the principal must be able to set the agent's continuation value following a deviation equal to her outside option. However if the principal must pay a transfer following a success on a costless task then the agent is receiving a per-period surplus on these tasks. In contrast, she is making a per-period loss when she correctly takes the costly action, and obtains nothing by deviating to the costless action. This implies that the agent can drive her continuation value above $U_0$, at least in the short run, and therefore cannot be persuaded to take the costly action. This will not be a problem if the costly action is only required on a single task, since the principal can avoid paying any transfers on the costless tasks until the costly task has been attempted. However if the costly action is required on more than one task, then for $\delta$ close to $\delta^*_\gamma$ then the fact that the relationship may break down in the furture may make it impossible to induce the agent to cooperate on a costly task at all.

It is worth noting that for certain values of $\delta$ and $p$, increasing the amount of information available (to the agent especially) can make cooperation more difficult to enforce. Let $TS(\gamma)$ denote the surplus generated when a fraction $\gamma$ of tasks are costly, and assume that $\delta = \delta^*_\gamma$. If the parties know only revenues, then they know that the surplus is at least $TS(\gamma)$, and so the expected value of the relationship is strictly greater than $TS(\gamma)$, for any $p \in (0, 1)$. In contrast, if the parties observe profits then they know that the value of the relationship is exactly $TS(\gamma)$. This uncertainty about the total value of the relationship may increase the transfers that the principal can credibly promise to the agent at the start of the game, especially for small $p$, and will increase the expected value to the agent of the relationship. Whilst this does not prevent the possibility that the relationship may be terminated on the equilibrium path following certain histories, for $\delta$ close to $\delta^*_\gamma$ it will make it easier to achieve cooperation at the start of the game.

A further implication of these results is that there may be parameter values for which an uninformed partnership is able to achieve the optimal level of cooperation, if they are lucky at the start of the game, but an informed partnership is deterred from attempting to cooperate at all. To see this, consider a case in which a fraction $\gamma$ of tasks require the costly action. Assume that $\delta$ is slightly greater than $\delta^*_\gamma$, and so it is just possible to induce the agent to supply the costly action when required in the game with full information. In the game with private information, if the costless tasks are all drawn first, and are successful, then it will be possible to obtain cooperation on the remaining costly tasks. (Alternatively, for smaller values of $p$ the relationship requires only that the agent is not required to work on too many costly tasks before encountering some costless tasks.) Moreover, for $p$ small the expected surplus of the relationship is such that the principal and the agent are both optimistic about the value of the relationship, and therefore willing to forego their outside option at the start of the game in order to attempt to build the relationship. However, if the realised value of $\gamma$ is $\gamma >> p$ then this may deter other partnerships from attempting to establish a relationship at all. We saw above that there is a positive probability of reaching a history in which the agent prefers to deviate to a costless action, expecting to receive a short-run surplus on certain tasks before terminating the relationship. The possibility of reaching such a history may reduce expected surplus from the relationship to a level at which the principal cannot credibly promise to pay the agent the transfer needed to ensure cooperation; knowing this, the parties may never attempt to enter a partnership.

## 1.7 Conclusion

This paper presents a model of relational contracts in which there is initally asymmetric information regarding the mapping from actions to output on a series of tasks, and therefore the principal cannot easily monitor the agent's actions. As the game goes on, the principal may be able to learn from past successes on each task, and thereby monitor the agent's actions on that task in the future. However, given these difficulties in monitoring and the initial uncertainty about the total value of the relationship it may not be possible to achieve full cooperation. The paper shows that there will exist regions of the parameter space and discount factors for which cooperation can be achieved in the game with full information, but cannot be assured in the game with private information. Moreover, whether or not it is possible to eventually produce at the maximal level may depend critically on the order in which tasks are first required, and on whether or not early attempts at each task are successful.

I go on to consider whether relaxing the assumptions on the equilibrium can increase the likelihood that the principal and the agent are able to maintain the relationship at an efficient level. In particular, the paper shows that allowing the agent a probationary period in which she is not expected to work on all tasks can alleviate some of the problems associated with uncertainty about the total surplus from the relationship, but cannot resolve all the difficulties of monitoring.

I argue that these results may partially explain why we see large differences in performance amongst firms in similar industries: even firms that share the same set of tasks and costs may end up performing at very different levels if the order in which they are required to undertake these tasks varies, and if some firms have more success on early tasks than others. The paper goes on to ask whether we should expect these differences in performance to be eliminated over time: if a poorly performing firm can see that an identical enterprise is functioning at a much higher level, they should be able to infer something about the value of the relationship, which may in turn make it easier to provide the correct incentives. I find that even learning the total surplus from the relationship with uncertainty cannot resolve all the incentive issues, and therefore we may still expect to see inefficient termination on the equilibrium path, and hence differences in performance, for certain values of $\delta$.

## 1.8 Appendix

### 1.8.1 Proof of Proposition 3

*Proposition 3: There exist values of $\alpha_i$, $\gamma_{-i}$ such that for $\delta \in [\delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}, \tilde{\delta}_{\gamma_{-i}(1-\alpha_i)+\alpha_i})$ there will be inefficiently low performance in equilibrium even though it is common knowledge that there is sufficient surplus to sustain the relationship at the optimal level.*

This result states that there exist values of $\delta \geq \delta^*_{\gamma_{-i}(1-\alpha_i)+\alpha_i}$ for which it is either impossible to induce the agent to take a costly action when appropriate, or doing so requires the threat of inefficient termination on the equilibrium path. This occurs for values of $\delta$ for which it is common knowledge that there is sufficient surplus to sustain the relationship.

*Proof.* In order for the agent to be induced to undertake the costly action, it must be the case that:

$$-c + \eta\tau_i(h^t, a_c) + \delta E[U(h^t, a_c)|a_i = a_c] \geq 0 + \delta \max\{E[U(h^t, a_0)|a_i = a_c], U_0\} \tag{1.9}$$

The bonus payments promised by the principal must also be credible, so

$$-\tau_i(h^t, a_c) + \delta E\Pi[h^t, a_c|a_i = a_c] \geq \Pi_0$$

where it is worth noting that the principal is only ever required to pay a bonus once the task is successful, meaning that the mapping from tasks to actions and the surplus of the relationship is now common knowledge. T We know that at $\delta^*_{\gamma_i(1-\alpha_i)+\alpha_i}$ this constraint holds with equality when the continuation value following a deviation is $U_0$. Therefore in order to ensure cooperation in the game with private information we need:

$$E[U(h^t, a_0)|a_i = a_c], U_0 \leq U_0 \tag{1.10}$$

However the principal cannot distinguish between a deviation to the costless action, and the agent correctly taking the costless action without success. In the latter case, the continuation surplus $E[U(h^t, a_0)|a_i = a_0]$ needs to be sufficient to ensure the agent supplies effort on costly tasks in the future:

$$-c + \eta \min\{\tau_j|a_j = a_c\} + \delta E[U(h^t, a_0)|a_i = a_0] \geq \delta U_0 \tag{1.11}$$

Therefore in order to reach the same level of performance as in the game with full information, we need equations (1.10) and (1.11) to hold at $\delta = \delta^*_{\gamma_i(1-\alpha_i)+\alpha_i}$. I will show that for $\alpha_i = 0$ these equations cannot hold for any $\gamma_{-i} > 0$, and therefore, through continuity in $\alpha$, there exists an open interval $[0, \hat{\alpha}_i)$ for which it is not possible to obtain the level of cooperation possible in the game with full information.

If $\alpha_i = 0$, these equations will hold with equality by setting $\tau_{i|a_i=a_c} = \frac{\delta(1-\delta)U_0+c}{\eta}$ and

$\tau_{i|a_i=a_0} = \frac{\delta(1-\delta)U_0}{\eta}$. This is possible as long as these transfer payments are credible for the principal. This requires that:

$$-\tau_{i|a_i=a_c} + \frac{\delta}{1-\delta}(\eta - \delta(1-\delta)U_0 - \gamma_{-i}c) \geq \delta\Pi_0$$

but this expression will not hold at $\delta^*_{\gamma_i(1-\alpha_i)+\alpha_i}$ for any $\gamma_{-i} > 0$. (This is the same argument made in Proposition 1: the set of values of $\delta$ under which cooperation is possible is maximised by paying the same transfer on all tasks.) For $\alpha_i > 0$ we must have $\delta E[U(h^t, a_0)|a_i = a_0] > \delta E[U(h^t, a_0)|a_i = a_c]$ since the agent is losing any bonus payments on task $i$. Nonetheless since this difference $\delta E[U(h^t, a_0)|a_i = a_0] - \delta E[U(h^t, a_0)|a_i = a_c]$ is continuous in $\alpha_i$, the necessary constraints cannot be met at $\delta = \delta^*_{\gamma_i(1-\alpha_i)+\alpha_i}$ for $\alpha \in [0, \hat{\alpha}_i)$.

We can also show that using the threat of termination may increase the set of values of $\delta$ for which it is possible to sustain cooperation. Assume that if the agent takes a costless action and is unsuccessful, then the principal can commit to terminating the relationship with probability $\xi(h^t)$. In this case cooperation requires that:

$$\begin{aligned} -c + \eta\tau_i(h^t, a_c) + \delta E[U(h^t, a_c)|a_i = a_c] \geq \\ (1 - \xi(h^t))\delta \max\{E[U(h^t, a_0)|a_i = a_c], U_0\} + \xi(h^t)\delta U_0 \end{aligned} \tag{1.12}$$

Clearly, by setting $\xi(h^t) = 1$ this incentive constraint reduces to that in the game with full information, ensuring that the agent will take the costly action if appropriate. The trade-off here is that since $\eta < 1$ there is a positive probability that the relationship will be terminated inefficiently on the equilibrium path; moreover this possibility reduces expected surplus earlier in the game, potentially making it more difficult to obtain effort on costly tasks. The resolution of this trade-off will depend on $p$ and $\eta$: for $\eta$ high, the expected loss of surplus due to inefficient termination is small; meanwhile for $p$ large, the possibility of being unable to obtain effort on a costly task carries more weight that the risk of inefficient termination.

□

### 1.8.2 Proof of Proposition 4

*Proof.* In order to ensure that the agent is willing to cooperate on a costly task $i$ that has not previously been attempted, given a history $h^t$, we require that:

$$-c + \eta\tau(a_c, h^t) + \delta E[U(a_c, h^t)|a_i = a_c] \geq \max\{\delta U_0, 0 + \delta E[U(a_0, h^t)|a_i = a_c]\}$$

and moreover that the principal must be willing to pay the transfer, if appropriate

$$-\tau(a_c, h^t) + \delta E[\Pi(a_c, h^t)|a_i = a_c] \geq 0$$

combining these expressions, given that the agent's continuation value following a deviation must be bounded below by $U_0$ we require that

$$c + \delta U_0 \leq \delta E[U(a_c, h^t)|a_i = a_c] + \eta\delta E[\Pi(a_c, h^t)|a_i = a_c] < \delta E[TS|a_i = a_c, h^t]$$

We now want to place a bound on this expected total surplus, conditional on the history of the game. I will show that following any history, if at time $t$ a fraction $\alpha_t$ tasks are already known, of which a fraction $\gamma_t$ are costly then the expected surplus in the game with private information is bounded above by the expected surplus in the game with full information in which a total fraction $\alpha_t\gamma_t + (1-\alpha_t)p$ of tasks are costly. That is,

$$E[TS_{full\ info}|\gamma_t, \alpha_t] \leq \frac{1}{1-\delta}\left[\eta - (\alpha_t\gamma_t + (1-\alpha_t)p)c\right]$$

We can do this inductively, starting with the case in which only 1 task remains unknown. Assume that this final task will occur in any period with probability $\alpha_i$, and that a fraction $\gamma_{-i}$ of the previously attempted tasks are known to be costly. The expected surplus given that only this task remains unknown is

$$E[TS|\gamma_{-i}, \alpha_i] = \frac{1}{1-\alpha_i\delta}\left[(1-\alpha_i)(\eta - \gamma_{-i}c) + \alpha_i(1-p)\left[\eta + \frac{\delta}{1-\delta}(\eta - \gamma_{-i}(1-\alpha_i)c)\right]\right] + \frac{1}{1-\alpha_i\delta}\left[\alpha_i p\left[(1-\mathbb{I}_{quit})\left[\eta - c + \frac{\delta}{1-\delta}(\eta - (\gamma_{-i}(1-\alpha_i) + \alpha_i)c)\right] + \mathbb{I}_{quit}\delta U_0\right]\right]$$

If task $i$ is costless, then if the relationship has survived until this point then it will not be terminated now. If not, the relationship may survive with cooperation on the costly task, or it may be terminated. If $\delta$ and $\gamma_{-i}$ are such that the relationship will survive then this expression simplifies to

$$E[TS|h^t] = \frac{1}{1-\delta}\left[\eta - (\gamma_{-i}(1-\alpha_i) + p\alpha_i)c\right]$$

This is exactly the expected surplus that would be generated in the game with full information if a fraction $\gamma_{-i} + p\alpha_i$ of tasks were known to be costly. That is if it were common knowledge that there were no risk of termination (for any reason), then it would be possible to achieve cooperation in the game with private information for some expected $\gamma$ whenever it is possible to achieve it in the game with full information and known $\gamma$. If, however, the relationship will terminate with positive probability if the task is costly then the expected surplus will be strictly less than that under full information.

Having established a bound for the expected surplus on the last task, consider some arbitrary point in the game. Assume that a fraction $\alpha_k$ of tasks are known. At time $t+1$, either a new task will be required, increasing the share of known tasks to $\alpha_{k+1}$, or one of the tasks already learned will be repeated. If a new task is drawn and it is costly, then the fraction of tasks known to be costly will increase to $\gamma_{k+1}(a_c)\alpha_{k+1} = \gamma_k\alpha_k + (\alpha_{k+1} - \alpha_k)$; if costless, then $\gamma_{k+1}(a_0)\alpha_{k+1} = \gamma_k\alpha_k$. We can therefore write the expected surplus as

$$E[TS|\gamma_k, \alpha_k] = \frac{1}{1-\alpha_k\delta}\left[\alpha_k(\eta - \gamma_k c) + (1-\alpha_k)(1-p)\left[\eta + E[TS|\gamma_{k+1}(a_0), \alpha_{k+1}]\right]\right] +$$
$$+\frac{1}{1-\alpha_i\delta}\left[(1-\alpha_k)(1-p)\left[(1-\mathbb{I}_{quit})\left[\eta - c + \delta E[TS|\gamma_{k+1}(a_c), \alpha_{k+1}]\right] + \mathbb{I}_{quit}\delta U_0\right]\right]$$

We know that the continuation surplus from quitting is strictly less than that from remaining in the relationship. From the inductive assumption, therefore,

$$E[TS|\gamma_k, \alpha_k] \leq \frac{1}{1-\alpha_k\delta}\left[(1-\alpha_k)p\left[\eta - c + \frac{\delta}{1-\delta}\left[\eta - (\gamma_k\alpha_k + \alpha_{k+1} - \alpha_k + (1-\alpha_{k+1})p)c\right]\right]\right]$$
$$\alpha_k(\eta - \gamma_k c) + (1-\alpha_k)(1-p)\left[\eta + \frac{\delta}{1-\delta}\left[\eta - (\gamma_k\alpha_k + (1-\alpha_{k+1})p)c\right]\right]$$
$$= \frac{1}{1-\delta}\left(\eta - (\alpha_k\gamma_k + (1-\alpha_k)p)c\right)$$

Therefore, following any history,

$$E[TS|\gamma_t, \alpha_t] \leq \frac{1}{1-\delta}\left[\eta - (\alpha_t\gamma_t + (1-\alpha_t)p)c\right]$$

and so we have an upper bound for the surplus in the game with private information. However, we also know that for $\delta < \delta_1^*$ and $p > 0$ there is a positive probability placed on histories for which the expected surplus is strictly less than this. This upper bound is constructed by taking the continuation value if the agent remains in the relationship rather than quitting. However for any $\delta < \delta_1^*$, if all tasks are revealed to be costly then the relationship will be terminated. For any $p > 0$ this occurs with positive probability, and so the expected surplus in the game with private information cannot attain this upper bound. $\square$

### 1.8.3 Details from the examples in section 5

*Obtaining costly effort on the second task when the first is costless*

For any $\delta \geq \delta_{1/2}^*$ it is easy to ensure that the agent is willing to cooperate should it require the costly action. The agent will take the costly action on task 2, knowing that task 1 requires the costless action, as long as:

$$-c+\eta\tau(a_c, a_1 = a_0)+\delta E[U(a_c)|a_1 = a_0, a_2 = a_c] \geq \max\{0+\delta E[U(a_0)|a_1 = a_0, a_2 = a_c], U_0\}$$

where $E[U(a)|a_1 = a_0, a_2 = a_c]$ denotes the expected continuation surplus for the agent, conditional on the agent taking the action $a$, given that the actions required are $(a_0, a_c)$. We also require the transfer to be incentive compatible, bearing in mind that in the event the principal is required to pay a transfer, it must be that the task has been successful and therefore the correct actions are common knowledge:

$$-\tau(a_2 = a_c, a_1 = a_0) + \delta E[\Pi | a_1 = a_0, a_2 = a_c] \geq 0$$

where $E[\Pi | a_1 = a_0, a_2 = a_c]$ if the principal's expected continuation surplus conditional on the tasks requiring actions $(a_0, a_c)$.

As long as $\delta E[U(a_0)|a_2 = a_c, a_1 = a_0] \leq U_0$, these constraints will be satisfied they are satisfied in game with full information, that is, whenever $\delta \geq \delta^*_{1/2}$. We need only check, therefore, that the principal can set $\delta E[U(a_0)|a_2 = a_c, a_1 = a_0] \leq U_0$. In this example with two tasks, this will not be a problem: if both tasks were costless, the principal could ensure that the agent remained in the relationship by setting transfers such that $E[U|a_1 = a_2 = a_0] = \frac{\eta\tau_{00}}{1-\delta} \geq U_0$. In this case the expected surplus from *deviating* to the costless action must be strictly less that $U_0$, given that there is no chance of success on one of the tasks.

*Obtaining costly effort at the start of the game*

The agent will take a costly action on the first task if:

$$-c + \eta\tau(a_1 = a_c) + \delta E[U(a_c)|a_1 = a_c] \geq 0 + \delta E[U(a_0)|a_1 = a_c]$$

and the principal is willing to pay the transfer:

$$-\tau(a_1 = a_c) + \delta E[\Pi | a_1 = a_c] \geq 0$$

We can place bounds on these continuation values: since the agent has the option of severing the relationship altogether, her continuation value following a deviation cannot be less than her outside option. The agent's incentive constraint can therefore be rewritten as:

$$\begin{aligned}
c + \delta U_0 &\leq \eta\tau(a_1 = a_c) + \delta E[U(a_c)|a_1 = a_c] \\
&\leq \eta\delta E[\Pi|a_1 = a_c] + \delta E[U(a_0)|a_1 = a_c] \\
&< \delta E[\Pi + U] = \frac{\delta}{2-\delta}\left[\eta - c + (1-p)\left[\eta + \frac{\delta}{1-\delta}(\eta - \frac{c}{2})\right] + p\delta U_0\right]
\end{aligned}$$

This expression for total surplus conditional on one costly task being drawn takes into account the expected time until the agent will draw, and learn, the second task (after which the relationship will either be terminated, or proceed as in the full information case), and the expected surplus generated by working on a single task in the interim. We can compare this expected surplus when, in expectation, $\frac{1+p}{2}$ tasks are costly to the surplus generated in the game with full information in which a fraction $\frac{1+p}{2}$ of tasks are known to be costly:

$$E[TS_{private\ info}|a_c] - E[TS_{full\ info}|\gamma = (1+p)/2] = \frac{[\delta(1-\delta)U_0 - (\eta - c)]}{(2-\delta)(1-\delta)}$$

This expression is strictly negative for all $p > 0$, and $\delta \in (0,1)$. For $p = 0$, there is no uncertainty about the cost of the remaining task, and so the game with full information and the game with private information generate the same surplus, and it will be possible to achieve cooperation on the costly task under the same conditions, that is, whenever $\delta \geq \delta^*_{1/2}$. As $p$ increases, the game with private information generates strictly less surplus, and therefore cooperation will only be possible for higher values of $\delta$, and in particular, cooperation will no longer be possible at $\delta = \delta^*_{1/2}$, for any $p > 0$. As $p$ approaches 1, it will not be possible to achieve cooperation on a costly task at the start of the game even for $\delta$ arbitrarily close to $\delta^*_1$.

*Equilibria with probation: ensuring the the agent remains in the relationship*

If $\delta \in [\delta^0_{1/2}, \delta^*_{1/2})$ then if the agent learns that exactly one task requires the costless action, then she will be willing to remain in the relationship, working only on that task. Therefore if the first task to be drawn requires the costless action, the relationship is secure. Otherwise, if the costly task is drawn then her expected surplus from the relationship will decrease, and she may quit. To remain in the relationship, we require that her expected surplus exceed the outside option:

$$EU_{partial}(p, \delta|a_1 = a_c) = \frac{\delta}{2-\delta}\left[p\delta U_0 + (1-p)\left[\eta\tau_0 + \frac{\delta}{1-\delta}\frac{\eta\tau_0}{2}\right]\right] \geq U_0$$

This condition cannot be satisfied for $p = 1$, and at $\delta = \delta^*_{1/2}$, given the maximum transfer that the principal can commit to paying, this expression holds only if $p = 0$. Therefore we can find values of $\delta^0(p) \in (\delta^0_{1/2}, \delta^*_{1/2})$ such that the agent will only remain in the relationship after drawing a costly task if $\delta \geq \delta^0(p)$. Given that the agent's expected surplus is decreasing in $p$, $\delta^0(p)$ is increasing in $p$, with $\delta^0(0) = \delta^0_{1/2}$, and $\delta^0(p') = \delta^*_{1/2}$ for some $p' < 1$.

The other intermediate case to consider is $\delta \in [\delta^*_{1/2}, \delta^*_1)$, in which it is possible to support cooperation on a single costly task. The agent will remain in the relationship if and only if

$$EU_{full}(p, \delta|a_1 = a_c) = \frac{\delta}{2-\delta}\left[p\delta U_0 + (1-p)\left[\eta\tau_{0,c} + \frac{\delta}{1-\delta}\left[\eta\tau_{0,c} - \frac{c}{2}\right]\right]\right] \geq U_0$$

As above, at $p = 1$ the outside option strictly exceeds the expected surplus from remaining in the relationship, and so for sufficiently large $p$ it will not be possible to convince the agent to remain in the relationship, even for $\delta$ very close to $\delta^*_1$. However at $\delta = \delta^*_{1/2}$ the agent will be willing to remain in the relationship for any $p \in [0, p'']$, where $p'' > p'$. We can see this by comparing the surplus generated at $\delta = \delta^*_{1/2}$ by working only on the costless task, to the surplus generated by working on both tasks. Full cooperation generates a strictly higher surplus for the agent than partial cooperation, and therefore if $EU_{partial}(p', \delta^*_{1/2}|a_1 = a_c) = U_0$, $EU_{full}(p', \delta^*_{1/2}|a_1 = a_c) > U_0$, implying that $EU_{full}(p'', \delta^*_{1/2}|a_1 = a_c) = U_0$ for some $p'' > p'$.

*Example with three tasks: incentives on the third task*

We know that the agent can be induced to take costly effort on a task $i$ if:

$$E[U(h^t, a_0)|a_i = a_c] \leq U_0$$

and so the continuation value from deviating is less than the outside option. The difference in the agent's expected utility from deviating to the costless action as opposed to correctly taking this action comes from the loss of expected transfers on this task in the future:

$$(1-\delta)E[U(h^t, a_0)|a_i = a_c] = (1-\delta)E[U(h^t, a_0)|a_i = a_0] - \alpha_i \eta \tau_i(h^t)$$

Assume that a task $j$ has been verified to be costly. In order to provide the necessary incentives on this task, the principal must set the transfers and the continuation surplus such that:

$$-c + \eta\tau_j(h^t, a_c) + \delta E[U(h^t, a_0)|a_i = a_0] \geq \delta U_0 \tag{1.13}$$

For $\alpha_i$ large, the fact that deviation to the costless task cannot possibly generate a success (and so the agent forgoes any transfers) may be enough to reduce the agent's continuation surplus following a deviation, whilst still generating enough surplus to support cooperation when the agent correctly takes the costless action. If Task $C$ is encountered last, then since $\alpha_C = 0.1$, this will not be the case here. However this will be sufficient to provide the incentives on Task A, where $\alpha_A = 0.45$. Alternatively, as $\delta$ increases the principal can provide the necessary incentives on Task A by promising a larger bonus on that task alone and reducing transfers on other tasks: increasing $\tau_A$ allows the principal to reduce $E[U(h^t, a_0)|a(C) = a_0]$, and thus to set $E[U(h^t, a_0|a(C) = a_c] \leq U_0$. At $\delta$ close to $\delta^*_{0.55}$ this will not be possible, and so the principal will be unable to obtain costly effort on this task.

*Example with three tasks: incentives on earlier tasks*

In this section I construct an upper bound for the value of the discount factor required to support cooperation on an earlier task, when there is still uncertainty about the costs of the remaining tasks.

Assume that one task $j$ has been verified as costly, and that this task has weight $\alpha_j$. Now the agent is required to take the costly action on task $i$, with weight $\alpha_i$. If the agent is to take the costly action, we require that:

$$-c + \eta\tau_i(h^t, a_c) + \delta E[U(a_c)|a_i = a_c, h^t] \geq 0 + \delta E[U(a_0)|a_i = a_c, h^t]$$

and the principal must be willing to pay the transfer:

$$-\tau_i(h^t, a_i = a_c) + \delta E[\Pi(a_c)|h^t, a_= a_c] \geq 0$$

Bounds can be placed on these continuation values: we must have $E[U(a_0)|a_i = a_c, h^t] \geq U_0$ since the agent can always choose to quit, in which case her continuation payoff becomes the outside option. We can then rewrite the agent's constraint as

$$\begin{aligned} c + \delta U_0 &\leq \eta\tau_i(h^t, a_i = a_c) + \delta E[U(a_c)|a_i = a_c, h^t] \\ &\leq \eta\delta E[\Pi(a_c)|h^t, a_i = a_c] + \delta E[U(a_c)|a_i = a_c, h^t] \\ &< \delta E[U + \Pi|a_i = a_c, h^t] \end{aligned}$$

We can then find an expression for the expected total value of the relationship as a function of $p$, given that the relationship will be severed should the 3rd task turn out to be costly.

$$\begin{aligned} E[U + \Pi|h^t] = & \frac{\delta}{1 - (\alpha_i + \alpha_j)\delta} [(\alpha_i + \alpha_j)(\eta - c) + (1 - \alpha_i - \alpha_j)p\delta U_0] + \\ & \frac{\delta}{1 - (\alpha_i + \alpha_j)\delta} \left[(1 - \alpha_i - \alpha_j)(1 - p) \left[\eta + \frac{\delta}{1 - \delta}(\eta - (\alpha_i + \alpha_j)c)\right]\right] \end{aligned}$$

Since $\delta < \delta_1^*$ we know that there is insufficient surplus to sustain cooperation on all tasks. The expected fraction of tasks that require the costly action, conditional on two tasks requiring the costly action is $(\alpha_i + \alpha_j + (1 - \alpha_i - \alpha_j)p)$. Let $\delta^*_{(\alpha_i+\alpha_j+(1-\alpha_i-\alpha_j)p)}$ be the discount factor necessary to sustain cooperation on all tasks in the game with full information, if a fraction $(\alpha_i + \alpha_j + (1 - \alpha_i - \alpha_j)p)$ are known to be costly. I show that in the game with private information the discount factor $\delta(a_c, p)$ required to induce cooperation on the second costly task is strictly greater than $\delta^*_{(\alpha_i+\alpha_j+(1-\alpha_i-\alpha_j)p)}$, for all $p \in (0, 1]$.

Cooperation on the second task requires: $E[U + \Pi|h^t] \geq c + U_0$ In the game with full information, cooperation on a fraction $(\alpha_i + \alpha_j + (1 - \alpha_i - \alpha_j)p)$ of costly tasks will be possible if

$$\frac{\delta}{1 - \delta} [\eta - c + (1 - (\alpha_i + \alpha_j))c] \geq c + U_0 \tag{1.14}$$

This latter expression will hold with equality for $\delta = \delta^*_{(\alpha_i+\alpha_j+(1-\alpha_i-\alpha_j)p)}$. In order to induce cooperation in the case with private information at $\delta = \delta^*_{(\alpha_i+\alpha_j+(1-\alpha_i-\alpha_j)p)}$, it must therefore be the case that the expected surplus in the game with private information is greater than or equal to expected surplus in the game with full information:

$$E[U + \Pi|h^t] - \frac{\delta}{1 - \delta} [\eta - (\alpha_i + \alpha_j + (1 - \alpha_i - \alpha_j)p)c] \geq 0$$

This expression simplifies to

$$p(\delta(1-\delta)U_0 - (\eta - c)) \geq 0$$

However by assumption all tasks generate an expected surplus and so $\eta - c \geq U_0$. Therefore expression (1.14) cannot hold for any $p \in (0,1]$. This implies that for any $p > 0$, the discount factor required to induce the agent to take a costly action in the game with incomplete information where the probability of the third task being costly is $p$ is strictly greater than the discount factor required to induce cooperation in the game with full information in which a fraction $\alpha_i + \alpha_j + (1-\alpha_i-\alpha_j)p$ of tasks are known to require the costly action.

As a corollary to this result, we can consider $\delta$ arbitrarily to close $\delta_1^*$. At $\delta_1^*$, the discounted expected surplus from the relationship is such that cooperation can be sustained even if all tasks are costly, essentially by paying transfers on a task-by-task basis. However once uncertainty about the total value of the relationship is introduced cooperation will not be possible even arbitrarily close to $\delta_1^*$, for any $p > \hat{p} \in (0,1)$. Letting $p(\delta)$ define the inverse function $\delta^{-1}(p)$, as $\delta \to \delta_1^*$, $\hat{p}(\delta) \to \hat{p}(\delta_1^*) < 1$.

We can also consider what happens as $p$ approaches zero, and so both parties can be confident that the third task will not also require a costly action. In this case, we will have $\delta(p) \to \delta^*_{\alpha_i+\alpha_j}$ as $p \to 0$. Proof: we know that $\delta^*_{(\alpha_i+\alpha_j+(1-\alpha_i-\alpha_j)p)} \to \delta^*_{\alpha_i+\alpha_j}$ as $p \to 0$. Moreover, for any $p \in (0,1)$, we know that $\delta(p) > \delta^*_{(\alpha_i+\alpha_j+(1-\alpha_i-\alpha_j)p)}$. From the continuity of $\delta(p)$ and the fact that $\delta(0) = \delta^*_{\alpha_i+\alpha_j}$, the result follows.

We could find a similar bound for the case in which the first task verified is costless, and compare the expected surplus:

$$E[U + \Pi | a_i = a_c, a_j = a_c] \leq \frac{\delta}{1-\delta}\left[\eta - (\alpha_i + \alpha_j + (1-\alpha_i-\alpha_j)p)c\right]$$

$$E[U + \Pi | a_i = a_c, a_j = a_0] \leq \frac{\delta}{1-\delta}\left[\eta - (\alpha_j + (1-\alpha_i-\alpha_j)p)c\right]$$

From these two expressions we can see:

1. In either case the surplus is decreasing in $\alpha_i$, the costly task on which cooperation is required. Therefore in cases in which the total surplus of the game remains uncertain, it is easier to induce cooperation on costly tasks that carry little weight: this is because learning such a task is costly has a smaller effect on the expected surplus from the relationship in the future.

2. If the first task verified is costless, then the expected surplus is increasing in $\alpha_j$; if costly, then it is decreasing in $\alpha_i$. Attempting an "important" task early on is therefore more risky in terms of the survival of the relationship.

Now consider cooperation on task $j$ at the start of the game.

Without solving for the optimal transfers necessary to induce the agent to take the costly action at the start of the game, we can show this result by placing a lower bound on the discount factor needed to ensure cooperation. This requires calculating an upper bound for the expected surplus from the relationship, and comparing this to the value of the relationship in the game with full information. We will do this by finding an expression for the expected total surplus assuming that there is no inefficient termination on the equilibrium path, following an unsuccessful attempt at the costless action. The outcome of the game will depend on the subsequent tasks drawn and whether (and at what point) the game may terminate.

Assume that $\delta \geq \delta^*_{\alpha_j}$, and therefore there is enough surplus to support cooperation on this task under full information. We can also define $\delta^*_{\alpha_j+\alpha_i}$ where $i \neq j$, as the discount factors required to support cooperation if a second task $i$ is costly. However we know that the discount factor required to support cooperation on this task if there is uncertainty about the cost of the third task is strictly greater than this. Therefore define $\delta(p, \alpha_j, \alpha_i)$ as the discount factor required to sustain cooperation on task $i$, given that task $j$ has been verified and the third task remains unknown: for $\delta \geq \delta(p, \alpha_j, \alpha_i)$ cooperation can be sustained on a second costly task $i$; for $\delta < \delta(p, \alpha_j, \alpha_i)$ the relationship will end if another action is revealed to be costly before a costless task is found.

There are three cases to be considered: (1) effort can be sustained on a second costly task $i$, without having verified that the third task is costless; (2) effort can be sustained on a second costly task $i$ only if the remaining task has first been verified as costly; and (3) the agent will not supply effort on a second costly task.

Consider the first case ($\delta \geq \delta(p, \alpha_j, \alpha_i)$) and the subgame after that task has been verified. If, after learning that task $j$ is costly, task $i$ is also found to be costly:

$$
\begin{aligned}
E[TS|a_c, a_c, \delta \geq \delta(p, \alpha_j, \alpha_i)] \leq & \frac{1}{1-(\alpha_i+\alpha_j)\delta}\left[(\alpha_i+\alpha_j)\eta - c + (1-(\alpha_i+\alpha_j))p\delta U_0\right] \\
& + \frac{1}{1-(\alpha_i+\alpha_j)\delta}\left[(1-(\alpha_i+\alpha_j))(1-p)\left[\eta + \frac{\delta}{1-\delta}\left(\eta - (\alpha_i+\alpha_j)c\right)\right]\right]
\end{aligned}
$$

Otherwise, if task $j$ is costly but task $i$ costless:

$$
\begin{aligned}
E[TS|a_c, a_0, \delta \geq \delta(p, \alpha_j, \alpha_i)] \leq & \frac{p}{\delta}\left[\eta - c + \frac{\delta}{1-\delta}\left(\eta - (1-\alpha_i)\right)\right] \\
& + \frac{1}{1-(\alpha_i+\alpha_j)\delta}\left[(\alpha_i+\alpha_j)\eta - \alpha_j c + (1-(\alpha_i+\alpha_j))(1-p)\left[\eta + \frac{\delta}{1-\delta}\left(\eta - \alpha_j c\right)\right]\right]
\end{aligned}
$$

In this case the order in which the tasks are drawn is irrelevant. For $\delta \geq \delta(p, \alpha_j, \alpha_i)$, cooperation can be achieved if a second costly task is drawn before a costless task has been revealed, and so in this case the expected surplus is:

$$E[TS|a_c, \delta \geq \delta(p)] \leq$$
$$\frac{1}{1-\alpha_j\delta}[\eta - c + \sum_{i\neq j} \alpha_i \left((1-p)[\eta + \delta E[TS|a_c, a_i = a_0]] + p[\eta - c + \delta E[TS|a_c, a_i = a_c]]\right)]$$

Consider the second case, for which cooperation on a second costly task $i$ will be possible only if a costless task is encountered first: $\delta^*_{\alpha_j+\alpha_i} \geq \delta < \delta(p, \alpha_j, \alpha_i)$. The continuation surplus after drawing a costly task at the start of the game is:

$$E[TS|a_c, \delta \geq \delta^*_{\alpha_j+\alpha_i}] \leq \frac{1}{1-\alpha_j\delta}\left[\eta - c + (1-\alpha_j)(1-p)[\eta + \delta E[TS|a_c, a_0]] + (1-\alpha_j)p\delta U_0\right]$$

In the third case, for smaller values of $\delta < \delta^*_{\alpha_j+\min_{i\neq j}\{\alpha_i\}}$ the relationship can be sustained only if neither of the remaining tasks are costly. The continuation surplus having encountered two of the tasks is therefore bounded by:

$$E[TS|a_c, a_0, \delta < \delta^*_{\alpha_j+\min_{i\neq j}\{\alpha_i\}}] =$$
$$\frac{1}{1-(\alpha_i+\alpha_j)\delta}\left[(\alpha_i+\alpha_j)\eta - \alpha_j c + (1-(\alpha_i+\alpha_j))(1-p)\left[\eta + \frac{\delta}{1-\delta}(\eta - \alpha_j)\right] + p\delta U_0\right]$$

and the continuation at the start of the game, having drawn a single costly task, is:

$$E[TS|a_c, \delta < \delta(p, \alpha_j)] \leq \frac{1}{1-\alpha_j\delta}[\eta - c + \sum_{i\neq j} \alpha_i \left[(1-p)[\eta + \delta E[TS|a_c, a_0]] + (1-\alpha_j)p\delta U_0\right]]$$

In all three cases, the expected surplus is strictly less than the expected surplus in the game with full information in which a fraction $(\alpha_j + (1-\alpha_j)p)$ of tasks are known to be costly

$$TS_{full\ info} = \frac{\delta}{1-\delta}\left[\eta - (\alpha_j + (1-\alpha_j)p)c\right]$$

Perhaps the easiest way to see this is by considering $p = 0$ and $p = 1$. At $p = 0$, both parties know that the remaining tasks are costless and so the surplus in the game with private information is equal to that in the game with full information. However for $p = 1$ it is straightforward to confirm that the surplus is strictly higher in the game with full information. From the fact that $TS_{full\ info}$ is linear in $p$, and $TS_{private}$ is convex in $p$, it follows that $TS_{full\ info} - TS_{private} > 0$ for all $p > 0$.

Since the expected surplus conditional on drawing a costly task $j$ in the game with private information is strictly less than that in the game with full information in which a fraction of tasks $\alpha_j + (1-\alpha_j)p$ are known to be costly, this places a bound on the parameter values for which cooperation is possible. We can see that this surplus is decreasing in $\alpha_j$, and therefore it is more difficult to obtain cooperation on costly tasks which occur frequently, given that this leads to a greater decrease in the expected surplus of the relationship. These expressions are then used to calculate the values of $\delta$ for which it is possible to sustain cooperation on Tasks A, B and C at the start of the game.

We can also use these expressions to find the parameter values for which the agent can be persuaded to remain in the game and acquire more information in the equilibrium with probation. In this case the cost to the agent is the opportunity cost of foregoing her outside option, so discount factor required decreases. However it will not restore efficiency compared to the game with full information.

# Chapter 2

# Walking The Talk: Mission Statements as Cheap Talk

## 2.1 Introduction

Many firms publish mission statements or expressions of firm philosophy, professing ideals concerning their treatment of employees or their place in the community. These statements can range from banal phrases about fairness, to more specific pledges to avoid layoffs wherever possible, to offering training and career advancement to employees, and so on. Given that these claims are not legally binding, they may be regarded with skepticism by employees, in which case they serve little purpose for the firm.

Whilst the sentiments expressed in many mission statements may seem trite, there nonetheless seems to be a consensus within management that mission statements matter. Studies have suggested that up to 85% of European and North American firms publish mission statements (Desmidt and Heene, 2006), and that a majority of managers believe them to be among the most important management tools (Bain & Co. 2003, Iseri-Say et al, 2008). Moreover, some research has claimed a link between certain forms of mission statements and performance: mission statements expressing values or the firm's philosophy appear to be associated with better performance, an association that doesn't seem to exist for more specific mission statements detailing the firm's technology or market base (Pearce II and David 1987, Musek 2008, Dermol 2012). However, it has also been noted that most mission statements fail the criterion of "reasonable disagreement" and thereby may in fact say very little (Leuthesser and Kohli, 1997).

Despite this consensus within the management literature that mission statements matter, there have been few attempts to model how mission statements may affect performance. Their influence has been attributed to "a heightened sense of purpose" (Pearce II, 1982) or to the fact that they specify a common organizational goal or code of behaviour (Bart, Bontis, Taggar, 2001). The literature has also highlighted the potential downsides of mission statements: too generic and they may seem platitudinous, too specific and they may make adaptation more difficult. Moreover mission statements may be open to misinterpretation, and subsequently risk laying management open to charges of hypocrisy (Edmondson and Cha, 2006).

This paper asks whether it is possible for an employer to relay credible information through such messages, and then reap the benefits via increased effort or loyalty to the firm on the part of employees. I present a model in which mission statements function as a form of cheap talk from firms to their employees, during a two-period game. I assume that firms differ in their quality, and can use the mission statement to send a costless message about their type to an agent at the start of the first period. Based on this message, the agent chooses the level of effort that he is willing to supply to the firm. I will assume that the principal's profit function is such that if the game were to terminate here, then the only cheap-talk equilibrium will be a babbling equilibrium. At the start of the second period of the game, there will be some probability that the firm will be called upon to follow through on their mission statement. This will take the form of a signaling game, in which the firm can destroy surplus in order to convey information about its type to the agent. The agent then has an opportunity to update his beliefs, before choosing how much effort to supply to the firm in the second period. I will assume that the agent's choice of effort at this second stage depends both on his earlier effort, and on his belief about the principal's type. Therefore, by affecting his earlier choice of action the message sent in the mission statement can influence both the agent's actions later in the game, and the cost of signaling. This in turn will influence the initial messages that different types of firm wish to send.

This paper shows that the possibility that the firm may have to act within the constraints of the mission statement, or face the ramifications of being regarded as hypocritical, may allow an informative cheap-talk equilibrium to exist in a setting in which cheap-talk would otherwise be uninformative. If we interpret mission statements in this way, then stronger statements of values or stronger commitments to employees represent messages corresponding to better types, and may be successful in eliciting higher effort. This provides a possible explanation for the widespread use of mission statements, and the model also predicts that broader statements of values may indeed have a greater impact than mission statements pertaining to a particular aspect of the firm's work, by increasing the probability that a situation will arise in which the firm is required to follow through on their promises. This paper also shows that there may be a trade-off between the informativeness of cheap-talk, and the costs that firms may have to bear through signaling. Whilst both agents and, in expectation, firms prefer more information to be conveyed through messages, this will impose additional signaling costs on the firm and so the net effect on firm profits and total welfare is ambiguous. I go on to consider whether communication via mission statements may hinder the firm's ability to adapt to new information, and show that the possibility that adaptation may be required in the future reduces the informativeness of initial communication. These effects match concerns expressed in the management literature.

For an example of how expressions of firm values may influence worker behaviour, consider the joint venture between Toyota and GM at the NUMMI (New United Motor Manufacturing Co. Inc) plant in California in 1984. The plant had closed two years previously in circumstances described as "ongoing war" (Brown and Reich, 1989) between the management and the United Auto Workers (UAW). However two years after reopening, the plant had increased its productivity by 50% to a level comparable to Toyota's Japanese plants and above that of any other GM plant. In contrast, at the Van Nuys plant, which had a history as troubled as NUMMI's, there was no improvement in productivity amongst ongoing tension between workers and management.

Whilst some of this success at NUMMI compared to Van Nuys may be attributed to better screening and selection of workers, we may also consider the different commitments and promises made at each plant, and the way in which they influenced the workers' expectations from the company. For example, NUMMI's agreement with the UAW stated that NUMMI "has a responsibility ... to provide stable employment," but that layoffs could be compelled by "severe economic conditions that threaten the long-term financial viability of the Company", although only if other policies, including reductions to executive pay, had already been attempted. Whilst the agreements made at the Van Nuys plant used similar language in general, the term "layoffs" was replaced by "*indefinite* layoffs" and the list of measures to be taken before any layoffs did not include reductions to management compensation. These may seem like small differences, particularly since the language used is sufficiently vague to avoid legal ramifications, but management at the two plants did in fact respond differently to the economic downturn in 1987-1988. Whilst the Van Nuys plant placed workers first on a "mandatory extended holiday" and later on 50% time with rotating shifts, the NUMMI plant honoured its commitment to avoid layoffs of any type and instead brought its workers in for further training at full pay. These different responses in turn may have affected workers' attitudes towards the firm: NUMMI's policies seemed to generate greater loyalty from the workforce, whilst at Van Nuys there was another period of disagreement between the management and the union, which felt shut out of decision-making.

In the context of the model presented in this paper, the early promises made by the management at NUMMI and Van Nuys constitute different cheap talk messages. Whilst neither company made binding commitments, NUMMI made somewhat stronger statements regarding potential layoffs than Van Nuys, and when called upon to act on these statements, followed through. Importantly, however some of the effects on worker behaviour at NUMMI were seen before any such demonstrable commitment to their workers had been made. Interpreted in the light of this model, this fact pattern suggests some separation through the initial messages.

The paper proceeds as follows. The next section discusses the related literature in economics, before proceeding to the model in section 3. Section 4 discusses the equilibria of this game, and how the separation induced in the cheap-talk game may depend on the equilibrium specified in the signaling stage of the game. Section 5 presents an example, and discusses the welfare implications of the trade-off between more costly equilibria in the signaling stage versus more informative communication in the initial cheap-talk game. Section 6 extends the model to allow for the possibility of a change in the firm's type and a consequent need for adaptation, and shows that this reduces the scope for informative cheap talk. Section 7 concludes.

## 2.2 Related Literature

In addition to the literature in management discussed above, there is considerable literature within economics discussing the role of leadership and values in organizations. Various papers have analysed leadership and firm missions in the context of a signaling game. For example Hermalin (1998) models leadership by example by allowing managers to send costly signals or make a sacrifice in order to convey information about the state of the world, inducing employees to supply more effort in high-productivity states of the world. In the event it occurs, the signaling game presented in this paper plays a similar role, but this paper adds an earlier opportunity for communication and shows that these early cheap-talk promises may convey coarse information about the firm's type, as long as there is positive probability that "leadership by example" may be called upon in the future.

In a paper that deals more explicitly with firm "missions" Bolton, Brunnermeier and Veldkamp (2011) present a model in which leaders use the firm mission statement to both convey information and induce cooperation amongst followers. However managers know that they will receive more precise information in the future, and so face a trade-off between inducing coordination through a clearly defined but potentially incorrect mission, and waiting for more information at the risk of miscoordination in the short term. A similar dynamic is captured in the extension to this paper, in which I consider the possibility that the firm's type may change. Dewan and Myatt (2008) also view communication from management as a means of facilitating coordination, and argue that more value should be placed on clarity than accuracy. In the model presented here both the principal and the agent value more information, absent the cost of signaling, but I show that increased clarity is associated with greater expected costs for the firm and the net welfare effect is ambiguous.

We can also relate this paper to the literature on managerial vision, such as Rotemberg and Saloner (2000) and Van den Steen (2005). In these papers the manager's type represents a particular vision or bias. In Rotemberg and Saloner hiring a manager with a particular bias allows the firm to commit to an overarching strategy, knowing that the manager will act in such a way as to enforce this strategy. This makes it easier to elicit effort from employees working on certain projects. Van den Steen shows that management vision may have a sorting effect and attract similar employees, who are more willing to supply effort. In contrast to these papers, I focus on a single task and a single employee type, but explicitly model the extent to which the manager's type may be communicated to employees before either party takes any costly actions.

This paper is also related to the literature on cheap-talk games that originated with Crawford and Sobel (1984). The most closely related paper is perhaps Kartik (2009) which considers a cheap-talk game in which there are costs that arise from fabricating misleading information, and which increase with the magnitude of the lie. In this paper, these costs associated with sending misleading information are endogenised via the signaling game, and can be regarded as the costs of being found out, rather than the costs of lying. As in the Kartik paper, theses costs increase with the magnitude of the lie, through the difference between the agent's belief based on the cheap-talk, and her updated belief following the signal

## 2.3 Model

A principal of privately known type $\theta \in [\underline{\theta}, \bar{\theta}]$ seeks to elicit effort from an agent during two time periods. Payoffs to the principal and the agent depend on the principal's type, and on the actions taken by the agent in each period. The payoff functions are such that the agent is willing to supply a higher level of effort to a principal whom he believes is of higher quality, whereas the principal always prefers the agent to supply more effort. Payoffs are realized at the end of the game.

There are two potential methods by which the principal can communicate her type to the agent: at the start of the first period the principal may send a costless message $m \in \mathscr{M}$, and at the start of the second period, with probability $p$, the principal may send a costly signal $s \in \mathbb{R}$, which may be thought of as money burning. We will focus on fully-separating equilibria in the signaling game, meaning that if the firm has an opportunity to signal, then in doing so it will reveal its type. The equilibrium concept will be Perfect Bayesian Equilibrium and so at each stage of the game the agent chooses his actions based on the available information.

The timing of the game is set out more formally below:

First Period:

1. The principal's type $\theta \sim U[\underline{\theta}, \bar{\theta}]$ is drawn and privately observed by the principal.
2. The principal sends a message $m \in \mathscr{M}$ to the agent.
3. The agent chooses an action $a_1 \in \mathscr{A}_1 = \mathbb{R}$.

Second Period:

4. With probability $p$, the principal has the opportunity to send a signal $s \in \mathbb{R}$, at a cost $c(s) = s$. The agent observes whether or not the signal is available and, if applicable, observes the signal $s$.

5. The agent chooses an action $a_2 \in \mathscr{A}_2 = \mathbb{R}$.

6. Payoffs $\Pi(a_1, a_2, \theta) - s$ and $U(a_1, a_2, \theta)$ are received.

The following assumptions are made regarding the payoff functions. Assume that the agent's utility function is weakly convex in $a_1$ and $a_2$ and that $U(a_1, a_2, \theta)$ satisfies increasing differences in $a_1$ and $\theta$, and in $a_2$ and $\theta$, and therefore the optimal choice of effort in each period is increasing in the agent's belief about $\theta$. Specifically, $U(a_1, a_2, \theta)$ is such that if for beliefs $f(\theta)$ and $g(\theta)$ over the principal's type $g(.) > f(.)$ in a first-order stochastic dominance sense, then $a_i(g(.)) > a_i(f(.))$ for $i = 1, 2$. In addition, assume that $\frac{\partial^2 U}{\partial a_1 \partial a_2} \leq 0$, so the agent's optimal choice of effort in the second period is decreasing in the effort supplied in the first period.

This assumption that the choice of effort in the second period is decreasing in first-period effort is crucial for the informativeness of cheap-talk in this setting. If the agent's utility were additively separable in effort, then effort in the second period would depend only on the belief induced by the signal, if applicable, and not on whether this belief corresponded with the message sent at the start of the game. In this case the principal would not be punished for sending a misleading message, and so would have no incentive not to imitate the highest type, rendering any messages uninformative. However if the agent's choice of action after observing the signal also depends on his effort earlier in the game, then inducing high effort in the first period could be costly for the principal if he is later caught out.

One way of thinking about this inseparability of effort in the agent's payoff function is to assume that the agent's payoff is some function of profits and the total effort expended. That is, $U(a_1, a_2, \theta) = f(\Pi(a_1, a_2, \theta)) - c(a_1 + a_2)$, where $c(a_1 + a_2)$ is concave. Given that the cost of second period will depend on first period effort, if the agent supplies high effort in the first period based on a high belief about the principal's type, but subsequently learns that he has been misled, then he may greatly reduce effort in the second period. Alternatively, these payoffs can be thought of from a behavioural point of view: if the agent feels that he has been deceived by the principal into supplying excessive effort at the start of the game, then he may resent this and punish the principal by withdrawing effort in the second period.

I now turn to the employer's profit function $\Pi(a_1, a_2, \theta) - s$. Assume that profits are increasing and concave in $a_1$, $a_2$ and $\theta$. Let $a_i(\tilde{\theta})$ denote the agent's optimal choice of action in each period, assuming the agent believes with certainty that the principal's type is $\tilde{\theta}$. I make the following assumptions:

1. $\frac{d\Pi(a_1(\tilde{\theta}), a_2(a_1(\tilde{\theta}), \theta), \theta)}{d\theta} > 0$ for all $\tilde{\theta} \leq \theta$;

2. $\frac{\partial^2 \Pi(a_1(\tilde{\theta}), a_2(a_1(\tilde{\theta}), \theta), \theta)}{\partial\theta \partial a_i} > 0$ for $i = 1, 2$;

3. $\frac{d^2 \Pi(a_1(\tilde{\theta}), a_2(a_1(\tilde{\theta}), \tilde{\theta}), \theta)}{d\tilde{\theta}^2} < 0$.

The first assumption states that for all types of employer $\theta$ the payoffs to imitating a type $\tilde{\theta}$ at the start of the game are strictly increasing for all $\tilde{\theta} \leq \theta$, even if the agent will learn the principal's type after the signaling game. This implies that all types can gain by imitating some better type. This assumption also rules out profit functions that would give the principal an incentive to destroy her reputation at the start of the game, in the hope of benefiting from an improvement in the agent's beliefs and an increase in effort following the signaling game. An alternative way of ensuring this would be to place restrictions on the relative weights of $a_1$ and $a_2$ in the profit function, and a constraint on the extent to which the agent's choice of effort $a_2$ would depend on $a_1$. It follows from assumption 1 that a weaker type would also have an incentive to imitate a stronger type if their true type wouldn't be revealed in the signaling game.

The second assumption requires that the principal's profit function satisfy increasing differences in $a_1$ and $a_2$, i.e., that a higher type gains more from higher effort in each period. Given that both $a_1$ and $a_2$ are increasing in the agent's belief about $\theta$, this in turn implies that the gains to imitating a particular type $\tilde{\theta}$ are increasing in the principal's true type $\theta$. The third assumption states that the principal's payoffs are concave in $\tilde{\theta}$, and so the returns to imitating a higher type are diminishing in that type.

## 2.4 Equilibrium

First, I define the equilibrium concept that will be applied throughout this paper. For the analysis of the equilibrium, I will begin by considering the agent's actions, given his beliefs about the firm's type. The paper then turns to the principal's problem, first solving for the principal's actions should the signaling game occur and then proceeding to analyse the cheap-talk messages sent in equilibrium. The paper presents conditions under which an informative cheap-talk equilibrium will exist, and analyses how this varies with the probability of the signaling game occurring, and the costs of the signals sent in equilibrium. For the first part of this section I consider the minimally informative cheap-talk equilibrium, that in which two distinct messages are sent. I subsequently consider equilibria in which more than two messages may be sent, and relax the assumptions made concerning the equilibrium in the signaling game.

Throughout this paper I will focus on a pure-strategy Perfect Bayesian Equilibrium of the game: the receiver (the agent) maximises his expected utility conditional on his belief after observing the message, and, if applicable, the signal; the agent's beliefs are defined by Bayes' rule, where possible; and the sender (the principal) maximises expected utility given her type and the receiver's strategy.

Should the signaling continuation game occur, a pure strategy for the principal is a function $s : \Theta \times \mathscr{M} \to \mathbb{R}$, so $s(m, \theta)$ is the signal sent by a type $\theta$ that sent an earlier message $m$. At the cheap-talk stage of the game, the principal's strategy is a function $\phi : \Theta \to \mathscr{M}$, defining the message $\phi(\theta)$ sent by type $\theta$. Turning to the agent, he will maximise his expected utility given his beliefs. Let $f(\theta|m)$ denote the agent's belief conditional on observing the message $m$; $f(\theta|m,s)$ describes the agents belief following the signal, if available. If the signal does not occur, let the agent's second period beliefs be denoted by $f(\theta|m,\varnothing)$. In the first period his strategy is a function $a_1 : \mathscr{M} \to \mathbb{R}$, a mapping $a_1(m)$ from the message to the agent's actions. In the second period his strategy is a function $a_2 : \mathbb{R} \cup \varnothing \times \mathscr{A}_1 \to \mathbb{R}$, so that $a_2(a_1, s)$ is a function of the agent's beliefs and his first period action.

To begin the analysis of the equilibrium, I first consider the agent's problem, and solve by backward induction. In the second period, the agent will choose his optimal action $a_2$ as a function of his earlier action and his belief about $\theta$. We can divide this into two cases: the case in which the principal has had an opportunity to signal her type and thus the agent knows $\theta$ with certainty, and the case in which no signal is available, and therefore the agent will not update his beliefs, which will be based solely on the message. Considering the first case, if a principal of type $\theta$ sends a costly signal then the agent will infer her type and choose effort accordingly:

$$a_2^*(a_1(m), s) = \arg\max_{a_2} U(a_1^*(m), a_2, \theta).$$

If no signal is received, the agent chooses effort to maximise his expected utility. Since the probability of receiving the signal is independent of the principal's type, the agent cannot infer anything from the fact that he received no information, and has no reason to update his beliefs: therefore $f(\theta|m,\varnothing) = f(\theta|m)$. In this case second-period effort solves

$$a_2^*(a_1(m), m) = \arg\max_{a_2} \int_\Theta U(a_1^*(m), a_2, \theta) f(\theta|m) d\theta.$$

Given these second-period actions, and the fact that the signal will be available with probability $p$, the agent's action in the first period is:

$$a_1^* = \max_{a_1} p \int_\Theta U(a_1, a_2(a_1, \theta), \theta) f(\theta|m) d\theta + (1-p) \int_\Theta U(a_1, a_2(a_1, m), \theta) f(\theta|m) d\theta. \quad (2.1)$$

The utility function is such that $a_2^*$ is decreasing in $a_1^*$. In the absence of a signal if $m' > m \Leftrightarrow E[\theta|m'] > E[\theta|m]$, then both $a_1$ and $a_2$ are increasing in $m$.

Given these actions on the part of the agent, we can turn to the principal's decision problem. First consider the principal's behaviour if the signaling game should occur. We will focus on the fully-separating equilibrium in which the principal's choice of signal $s$ perfectly reveals his type. This requires that the signal $s(\theta, m)$ is such that:

$$\theta = \arg\max_{\tilde{\theta}} \left\{ \Pi(a_1^*(m), a_2^*(a_1^*(m), s(\tilde{\theta})), \theta) - s(\tilde{\theta}, m) \right\}.$$

Solving for the signal sent by a type $\theta$:

$$s(\theta, m) = s(\underline{\theta}(m)) + \int_{\underline{\theta}(m)}^{\theta} \frac{d}{d\tilde{\theta}} \Pi(a_1(m), a_2(a_1(m), s(\tilde{\theta}, m)), x)\Big|_{\tilde{\theta}=x} dx. \tag{2.2}$$

The signal sent by a type $\theta$ therefore depends on the lowest type $\underline{\theta}$, the minimum of the support, but not on the distribution over the support. If there exists a non-babbling cheap-talk equilibrium then the message $m$ sent at the beginning of the game can partition the type space into intervals. The lowest type $\hat{\theta}_i(m_i)$ in each interval will determine the signal sent by the remaining firms in that interval. For now, assume that the least-cost fully separating equilibrium is played over each interval induced by the cheap-talk game, i.e., $s(m_i, \hat{\theta}_i) = 0$. This implies that any type $\hat{\theta}$ that lies at the cutoff point between two messages must bear a strictly larger signaling cost from sending the lower message than the higher message. Later in the paper I will relax this assumption and show that allowing for more costly signaling equilibria can increase the informativeness of communication, but at the expense of greater inefficient signaling costs.

I now turn to the cheap-talk game, and look for the minimally informative non-babbling equilibrium in which messages are sent that partition the type space into two intervals. For simplicity of notation, let $\Pi(m, \theta) \equiv \Pi(a_1(m), a_2(a_1(m), m), \theta)$ denote the profits in the event that the signaling game does not occur, as a function of the message sent $m$ and the principal's type $\theta$. Let $\Pi(m, s_m, \theta) \equiv \Pi(a_1(m), a_2(a_1(m), s), \theta)$ denote the profits as a function of the signal, should it become available.

**Proposition 2.** *A non-babbling cheap-talk equilibrium that partitions the type-space into two intervals will exist in this game if there exists a type $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ that is indifferent between sending the lower message and being believed to be in the set $[\underline{\theta}, \hat{\theta}]$, and sending the higher message and being believed to be in the set $[\hat{\theta}, \bar{\theta}]$. The type $\hat{\theta}$ must satisfy:*

$$(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] = (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s(\hat{\theta}), \hat{\theta})]. \tag{2.3}$$

To show that this is a sufficient condition for a non-babbling cheap-talk equilibrium of this form to exist it is necessary to verify that the cutoff $\hat{\theta}$ defined by equation (2.3) does in fact partition the type-space into two, and so all types $\theta \geq \hat{\theta}$ must prefer to send the message $m_H$, and all types $\theta < \hat{\theta}$ the message $m_L$, given the agent's beliefs. In this case, the agent will respond to these messages by supplying strictly greater effort in the first period after receiving the higher message. It follows from the fact that the principal's payoff function satisfies increasing differences between $a_1$ and $\theta$, that higher types gain more from inducing a higher belief, and therefore if there is some type $\hat{\theta}$ that is indifferent between sending the message $m_L$ and the message $m_H$, all higher types must strictly prefer to send the higher message, all lower types the lower message. The details are in the appendix.

### 2.4.1 Increasing the probability that the signaling game is played

In this section I provide further characterisation of the cheap-talk equilibrium. The next section of the paper states two lemmas. The first shows that for a given cutoff $\hat{\theta}$, which will determine the agent's beliefs in the absence of other information, as $\hat{\theta}$ increases the difference in payoffs between sending the lower message versus the higher message increases. A corollary of this result is that if the lowest type $\underline{\theta}$ prefers pooling with the remaining types to announcing her true type, then an informative equilibrium cannot exist. This provides a simple necessary condition for the existence of a non-babbling equilibrium. The second lemma shows that if the type space is partitioned into two intervals $[\underline{\theta}, \hat{\theta})$ and $[\hat{\theta}, \bar{\theta}]$ then $\bar{\theta} - \hat{\theta} > \hat{\theta} - \underline{\theta}$. That is, given that all principals have an incentive to imitate a higher type, the high message will convey less information than the low message.

Following these results, I can then show that the informativeness of communication is increasing in $p$: both in the sense that if a non-babbling equilibrium exists for some $p$, it will also exist for all $p' > p$, and in the sense that the expected variance of $\theta$ conditional on the message is decreasing in $p$.

**Lemma 1.** *If*

$$(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] < (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s(\hat{\theta}), \hat{\theta})] \quad (2.4)$$

*for some $\hat{\theta}$, then this is also true for all $\hat{\theta}' > \hat{\theta}$.*

This result states that increasing the cutoff $\hat{\theta}$ increase the gains for that type of being believed to be in the higher interval. Note that since we are varying $\hat{\theta}$, this is an off-equilibrium result. The proof is in the appendix, but the intuition comes from the fact that both the agent's utility function and the principal's profit function satisfy increasing differences between $a_1$ and $\theta$. Whilst increasing the cutoff $\hat{\theta}$ increases the agent's beliefs conditional on either message, the increase in effort will be greater for the higher message, and in turn this generates greater profits for a higher type principal. Given this result, we can now state the following:

**Corollary 1.** *If*

$$(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] < (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s(\hat{\theta}), \hat{\theta})] \quad (2.5)$$

*for $\hat{\theta} = \underline{\theta}$, then a non-babbling cheap-talk equilibrium of this game does not exist.*

This is a direct result of the lemma above. This result means that if the weakest type preferred not to send a distinct message $m(\underline{\theta})$, even if it were only type to do so, then a non-babbling cheap-talk equilibrium cannot exist. Therefore a necessary condition for the existence of a non-babbling equilibrium is:

$$\Pi(a_1(\underline{\theta}), a_2(a_1, \underline{\theta}), \underline{\theta}) \geq (1-p)\Pi(a_1(E[\theta]), a_2(a_1, E[\theta]), \underline{\theta}) + p\Pi(a_1(E[\theta]), a_2(a_1, \underline{\theta}), \underline{\theta}). \quad (2.6)$$

This expression comes from the fact that the lowest type $\underline{\theta}$ must be able to reveal her type at zero cost, should the signaling game occur. It also follows that in looking for the cutoff $\hat{\theta}$ that makes the principal indifferent between sending the message $m_L$ and the message $m_H$, we must ensure that this cutoff $\hat{\theta} \geq \underline{\theta}$.

**Lemma 2.** *If a non-babbling cheap-talk equilibrium exists in which the type-space is partitioned into two, then it must be the case that the indifferent type $\hat{\theta} < \frac{\underline{\theta}+\bar{\theta}}{2}$.*

This lemma means that if the type-space is partitioned into two, the upper interval must exceed the lower interval, and therefore stronger messages convey less precise information. Intuitively, if we assume some indifferent type $\hat{\theta} > \frac{\underline{\theta}+\bar{\theta}}{2}$, then from our assumption of a uniform distribution over $[\underline{\theta}, \bar{\theta}]$, the difference between the agent's belief and the principal's true type will be greater if the principal sends the lower message than if he sends the higher message. Given that the profit function is concave in $\theta$, $\hat{\theta}$ must strictly prefer to induce the higher belief in this case, even without accounting for the additional cost of signaling that she would incur if she sent the lower message.

The two lemmas stated here allow us to analyse how the existence of a non-babbling cheap-talk equilibrium may depend on the probability of the signal becoming available, and to provide some comparative statics on the amount of information conveyed to the agent. Intuitively, since only a babbling cheap-talk equilibrium exists in the game without the signaling opportunity, and it is the constraints on the firm's signal that provide the traction for informative cheap-talk, we would expect a higher probability of the signal to be associated with more informative communication since it increases the costs associated with misrepresenting the firm's type.

**Proposition 3.** *If an informative cheap-talk equilibrium exists for some $p$, then an informative cheap-talk equilibrium will also exist for all $p' > p$, with $\hat{\theta}' > \hat{\theta}$. Moreover the information conveyed by the message is increasing in $p$, in the sense that $E[Var(\theta|m)]$ is decreasing in $p$.*

*Proof.* As stated above, the existence of an informative equilibrium in this game requires that there exists $\hat{\theta}$ such that:

$$(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s, \hat{\theta}) - s(m_L, \hat{\theta})] = (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s, \hat{\theta})]$$

which implies that:

$$p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] = (1-p)[\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta})].$$

Since we are focusing on profit functions for which cheap-talk cannot be informative in the absence of the signal, we know that the right-hand side of this expression is positive for all $\hat{\theta}$: firms of all types $\theta$ prefer to send the "better" message if there is no risk of being called upon to act on this message. Assume that this expression holds with equality for some $p$, $\hat{\theta}$. Now consider increasing $p$ to $p'$. In this case the left-hand side will exceed the right-hand side. From Lemma 1 we know that, for $\theta' > \hat{\theta}$

$$\Pi(m_H, s(\theta'), \theta') - \Pi(m_H, s(\hat{\theta}), \hat{\theta}) > \Pi(m_L, s(\theta'), \theta') - \Pi(m_L, s(\hat{\theta}), \hat{\theta})$$

and therefore $\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta})$ is decreasing in $\theta$. Therefore in order to have equation (1) hold with equality for $p' > p$, we must have $\hat{\theta}' > \hat{\theta}$. Since we must have $\hat{\theta}' < \frac{\theta+\bar{\theta}}{2}$, it cannot be the case that $\hat{\theta}'$ now exceeds $\bar{\theta}$, and so the non-babbling equilibrium will still exist.

Turning to the second part of this proposition, we know from above that $\hat{\theta} < 1/2$, and so increasing $p$ shifts the cutoff towards $1/2$, making the two partitions more equal. The expected variance of the agent's belief, conditional on the message sent, is $E[\mathrm{Var}\,(\theta|m)] = \hat{\theta}\frac{\hat{\theta}^2}{12} + (1-\hat{\theta})\frac{(1-\hat{\theta})^2}{12}$, which is decreasing in $\hat{\theta}$ for all $\hat{\theta} \in [0, 1/2)$. Whilst the decrease in uncertainty following the high message is offset by an increase in uncertainty following the lower message, the fact that the firm's type is more likely to lie in the higher interval ensures that the net effect is to reduce the agent's uncertainty. □

This is as one would expect. Given that informative communication is not possible if the firm is never called upon to follow up on their promises, increasing the chance that the signal is available and hence that the firm will face a cost of lying makes initial communication more informative. Moreover increasing the chance that the firm is caught misleading that agent and is punished by the reduction of effort in the second period reduces the principal's temptation to imitate a better type, thereby reducing the mass of types that wish to send a higher message. This reduces the expected variance in the principal's type conditional on the message sent, thus increasing the informativeness of the equilibrium. However since signals in this game may be interpreted as money-burning and are thus wasteful, the gains from improved communication may be offset by the increase in expected cost.

Considering this result in the context of mission statements, this suggests that a mission statement is more likely to be informative as the probability of having to follow up on the promises made therein increases. This might explain the somewhat counterintuitive suggestion from the management literature that broader mission statements are more effective than statements that relate to a particular aspect of the firm's work, such as their technology or marketing. Whilst statements regarding a specific area of the firm might be expected to convey more information, if there is little chance that an opportunity will arise for the firm to demonstrate their commitment within a particular area, workers will be less inclined to give them much weight. However a mission statement detailing the firm's core values or their commitments to all employees could have more credibility, by increasing the set of circumstances in which the firm might have to act on its promises, and thereby increasing $p$.

### 2.4.2 Equilibria in which more than two messages are sent

Thus far, the paper has focused on the minimally informative non-babbling cheap-talk equilibrium, that in which the messages sent in equilibrium succeed in credibly dividing the type-space into two partitions. I now turn to the more general case, and define conditions under which the type space may be divided into $n$ partitions, and show that the informativeness of communication is increasing in $p$, in the sense that if a cheap-talk equilibrium with $n$ partitions exists for some $p$, then a cheap-talk equilibrium with $n' \geq n$ partitions will also exist for all $p' > p$.

Assume that there are now $n$ partitions $\{\underline{\theta}, ..., \hat{\theta}_i, ..., \hat{\theta}_{n-1}, \bar{\theta}\}$, where types in the partition $[\hat{\theta}_{i-1}, \hat{\theta}_i)$ send a message $m \in \mathscr{M}_i$, where $\mathscr{M}_i \subset \mathscr{M}$, and $\mathscr{M}_i \cap \mathscr{M}_j = \varnothing$ for all $i \neq j$. By a similar argument to that made above, we can show that the type-space will indeed be partitioned in this way; that is, if some type $\theta$ sends a message $m \in \mathscr{M}_i$ that induces a belief $f(\theta|m)$, then any type $\theta' > \theta$ will send a message $m'$ that induces a belief $f(\theta|m') \geq f(\theta|m)$. This implies that stronger types send weakly better messages, and ensures that we do not have to consider equilibria in which, for example, the strongest and weakest types send the same message, but intermediate types another.

**Proposition 4.** *A non-babbling cheap-talk equilibrium in which the type-space is partitioned into n intervals will exist is there are a series of types $\{\hat{\theta}_i\}$ such that:*

$$(1-p)\Pi(m_i, \hat{\theta}_i) + p[\Pi(m_i, s(\hat{\theta}_i, m), \hat{\theta}_i) - s(m_i, \hat{\theta})] < (1-p)\Pi(m_{i+1}, \hat{\theta}_i) + p[\Pi(m_{i+1}, s(\hat{\theta}_i, m), \hat{\theta}_i)] \tag{2.7}$$

*for each $i \in 1, ..., n-1$, where*

1. *message $m_i$ induces the belief that $\theta \sim U[\hat{\theta}_{i-1}, \hat{\theta}_i)$; and*

2. *$\hat{\theta}_0 = \underline{\theta}$, $\hat{\theta}_n = \bar{\theta}$.*

*Moreover if an equilibrium with n partitions exists for some p, then for any $p' > p$, an equilibrium with $n' \geq n$ partitions will exist.*

This result is analogous to the case in which the type-space is partitioned into two. Firstly, the messages will partition the type-space in a monotonic way, with a series of intervals in which higher intervals correspond to "higher" messages, i.e., messages that generate beliefs over a higher range of types, and hence lead to higher first-period actions. From the fact that the profit function satisfies increasing differences in $a_1$ and $\theta$, if some type $\theta$ derives greater utility from sending the message $m_{i+1}$ rather than the message $m_i$, all types $\theta' > \theta$ must also prefer this stronger message.

The argument that the maximum number of informative messages that can be sent in equilibrium is increasing in $p$ also follows from arguments made above. Assume, for example, that for fixed $p$ there is a non-babbling equilibrium in which the type-space is partitioned into two. Now consider $p' > p$. From Proposition 2 above, the informative equilibrium will still exist and the cutoff $\hat{\theta}$ between the two messages will increase. Will it be possible to generate a more informative equilibrium in this case? From Corollary 1, if a more informative equilibrium exists, then it must be the case that the lowest type $\underline{\theta}$ prefers to announce its true type (and have the agent believe that message and ascribe $\Pr[\theta = \underline{\theta}|m] = 1$), to pooling with the types in $[\underline{\theta}, \hat{\theta}]$. That is,

$$\Pi(a_1(\underline{\theta}), a_2(\underline{\theta}), \underline{\theta}) > p\Pi(a_1(m_L), a_2(a_1(m_L), \underline{\theta}), \underline{\theta}) + (1-p)\Pi(a_1(m_L), a_2(a_1(m_L), m_L), \underline{\theta}).$$

The right-hand side of this expression is decreasing in $p$, since this increases the probability that the agent learns the firm's true type and reduces second-period effort accordingly. Therefore for sufficiently large $p$ the weakest type may prefer to send a distinct message, and the type-space can be partitioned into three. This argument extends to the general case in which there are $n$ messages, and therefore the number of partitions that can be achieved in the most informative cheap-talk equilibrium in weakly increasing in $p$.

Increasing the probability that a firm is called upon to act on their promises not only weakly increases the number of messages that may be sent in equilibrium, but it also increases the informativeness of such an equilibrium. This can occur through two routes: firstly, for a fixed number of intervals, increasing $p$ increases each of the cutoffs, reducing the expected variance of $\theta$ conditional on $m$, as above; secondly it may increase the number of messages sent, and the existence of an additional interval will reduce the size of each partition, and hence the variance of the principal's type conditional on the message. I address each of these in turn.

The first effect, that the expected variance falls, is analogous to the result of Proposition 2. Increasing $p$ increases each cutoff $\hat{\theta}_i$, which will make the partitions of the type-space more equal. The expected variance of the message conditional on the signal in this more general case is $E[\mathrm{Var}(\theta)|m] = \sum_{i=1}^{n} \hat{\theta}_i \frac{(\hat{\theta}_i - \hat{\theta}_{i-1})^2}{12}$ where $\hat{\theta}_0 = \underline{\theta}$ and $\hat{\theta}_n = \bar{\theta}$. Minimizing this expression implies that $(\hat{\theta}_{i+1} - \hat{\theta}_i)^2 = (\hat{\theta}_i - \hat{\theta}_{i-1})^2$ for all $i$, i.e., all intervals should be equal in size. Whilst this will not be attainable in this game, increasing $p$ will bring this expected variance closer to the minimum. As in the two message case, for lower messages the variance of the agent's belief may increase, but the ex ante expectation of this variance will decrease.

The second effect is somewhat stronger: if the number of partitions in a cheap-talk equilibrium increases from $n$ to $n+1$ then the cutoffs between partitions in the new case will be such that the expected variance conditional on $m$ decreases for every message sent, by any type $\theta$.

We can see this in an example. Consider a cheap-talk equilibrium in which 2 messages $(m_L, m_H)$ are sent in equilibrium, as in the previous section, where the indifferent type is $\hat{\theta}$. As shown above, if the lowest type $\underline{\theta}$ strictly prefers to announce her true type to sending the message $m_L$, then there will exist a cheap-talk equilibrium with 3 intervals. Assume that $p$ is such that this lowest type is indifferent, and that when the probability of the signaling game increases from $p$ to $p'$ this type now strictly prefers to send a distinct message. Let this lowest type $\underline{\theta}$ send a "new" message $m_0$; if this type strictly prefers to send this message then there is some type $\hat{\theta}_0 > \underline{\theta}$ that is now indifferent between sending the message $m_0$ and the message $m_L$ (given that the low message now induces different beliefs). We can now turn to the type $\hat{\theta}$, who was previously indifferent between the messages $m_L$ and $m_H$: since the very lowest types now send the message $m_0$ instead of $m_L$, the agent's belief about $\theta$ conditional on observing $m_L$ will improve, so the agent will supply more effort after receiving this message. This implies that the type $\hat{\theta}$ must now strictly prefer to send the message $m_L$, and that the indifferent type must be some $\hat{\theta}' > \hat{\theta}$.

Moreover we will also have $\hat{\theta}' - \hat{\theta} < \hat{\theta}_0 - \underline{\theta}$. This comes from increasing differences between $a_1$ and $\theta$. We know that in the equilibrium with two messages, $\underline{\theta}$ was indifferent between sending the message $m_0$ and the message $m_L$, and the type $\hat{\theta}$ was indifferent between the messages $m_L$ and $m_H$, given $a_1(m_L)$. We also know that if a third message is introduced the lower bound on the set of types that send the message $m_L$ must increase from $\underline{\theta}$ to $\hat{\theta}_0$, and so $a_1(m_L)$ must increase. From the fact that there are increasing differences between $a_1$ and $\theta$, this increase in effort generates a greater increase in profits for a better type. This implies that a smaller increase in $\hat{\theta}$ is required to restore indifference between the messages $m_L$ and $m_H$, given that this increase in $\hat{\theta}$ will also increase $a_1(m_H)$. Therefore the set of types sending the message $m_L$ will decrease in moving from the equilibrium with two messages to that with three messages.

The increase in the number of messages sent must also increase the amount of information conveyed by *any* message. We know that the set of types sending the message $m_L$ must decrease, whilst the increase in the cutoff $\hat{\theta}$ must also decrease the set of types sending the message $m_H$. This implies that those types that send the same message in the equilibrium with three types as in the equilibrium with two types must now lie in a smaller interval, and so their messages must convey more information. Meanwhile, those types that send a different message must send a lower message, i.e., a more informative message (from Lemma 2). Thus the variance of the agent's beliefs conditional on the message $\text{Var}(\theta|m)$ decreases for all messages. This is stronger than the previous result, that stated that increasing $p$ without increasing the number of messages sent in equilibrium increases the *expected* variance conditional on the message.

### 2.4.3 Increasing the cost of signaling

So far I have focused on the least-cost fully-separating equilibrium in the signaling continuation game. Whilst this may be justified as the unique equilibrium by applying the D1 refinement (Cho and Sobel 1990), it is worth asking whether allowing other, more costly, equilibria in the signaling game may affect the informativeness of communication. The least-cost separating equilibrium may moreover seem counterintuitive, in that the lowest type of firm in the highest interval reveals its type at zero cost, whereas firms of lower quality, which did not make such strong promises, still send costly signals. It might be reasonable in such a scenario for the employees to conclude that the firm is of the worst possible type, or that they had misinterpreted the initial message sent by the firm. Imagine for example that a firm in the lower interval had been misinterpreted by the agent and assumed to belong to a higher interval. In the equilibrium described above, where firms of very different types, which sent different messages, may nonetheless end up sending the same signal, the weaker firm would not be able to distinguish itself later in the game.

In this section of the paper I consider equilibria in which the signal sent by the lowest type $\hat{\theta}_i$ that sends the message $m_i$ is strictly positive. Let $s(\hat{\theta}_i, m_i)$ denote this signal. In this setting the agent cannot extort arbitrarily costly signals from the principal, although the ability to do so would allow the agent to guarantee that the principal always announces her type truthfully. The cost of the signal is constrained by the fact that the principal always has the option of refusing to pay, i.e., setting $s = 0$, and accepting that the agent will respond to this by assuming that the principal is the worst type. Therefore the largest signaling cost $s_{max}(\hat{\theta}_i)$ that can be imposed on a type $\hat{\theta}_i$ satisfies:

$$\Pi(a_1(m_i), a_2(a_1(m_i), \hat{\theta}_i), \hat{\theta}_i) - s_{max}(\hat{\theta}_i, m_i) = \Pi(a_1(m_i), a_2(a_1(m_i), \underline{\theta}), \hat{\theta}_i). \qquad (2.8)$$

The following proposition considers the two-type case and states that increasing the signaling cost in this way increases the amount of information communicated by the cheap-talk message. The same result will also apply for the case in which $n$ messages are sent.

**Proposition 5.** *The informativeness of communication is increasing in the cost of signaling, $s(m_H, \hat{\theta})$: if a non-babbling cheap-talk exists for some $s(\hat{\theta}, m_H) \in (0, s_{max}(\hat{\theta}_i, m_i)]$, it will also exist for all $\tilde{s}(\hat{\theta}', m_H) > s(\hat{\theta}, m_H)$, where $\hat{\theta}' > \hat{\theta}$.*

This proposition states that increasing the signaling cost for all types in the higher interval, by increasing the signal sent by the lowest type in that interval, makes cheap-talk more informative both by increasing the range of values $p$ for which such an equilibrium exists, and by increasing the cutoff $\hat{\theta}$ between messages (which as noted above, reduces the expected variance of the agent's beliefs following the message). The result that this cutoff cannot exceed $\frac{\underline{\theta}+\bar{\theta}}{2}$ still holds (see proof of Lemma 2 in the appendix), and so a non-babbling equilibrium must still exist.

*Proof.* Let $\hat{s}$ denote the signal sent by the indifferent type $\hat{\theta}$. The indifferent type $\hat{\theta}$ must satisfy

$$p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta}) - [s(m_L, \hat{\theta}) - \hat{s}] = (1-p)[\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta})].$$

Increasing $\hat{s}$ to $\hat{s}'$ increases the right-hand side of this expression. As above, increasing differences in the principal's profit function implies that $\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta})$ is increasing in $\hat{\theta}$, whereas $\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta})$ is decreasing in $\hat{\theta}$. Therefore if the cost of signaling increases to $\hat{s}' > \hat{s}$, the indifferent type must now be some $\hat{\theta}' > \hat{\theta}$. $\square$

Thus an equilibrium in which all firms that send a higher message must pay a more costly signal can increase the information available to the agent. This may be due to the fact that by increasing $\hat{s}$ it is now possible to obtain a non-babbling cheap-talk equilibrium for payoffs, or a value of $p$, under which such an equilibrium would not exist for $\hat{s} = 0$. Moreover, extending this argument to allow for a general number of partitions, increasing $\hat{s}$ may increase the maximum number of distinct messages that can be sent in equilibrium. In addition, and analogously to varying $p$, by shifting the cutoffs between messages, increasing $\hat{s}$ can reduce the expected variance of the agent's beliefs conditional on the message.

This increase in the signaling cost effectively increases the punishment imposed on the principal by lying: if a low type seeks to deviate by sending a higher message, she must bear a larger cost of maintaining this pretence should the signaling game occur. Whilst this generates more informative communication, it also increases the expected value of surplus destroyed through wasteful signaling. This might explain some of the criticisms of mission statements found in the management literature: for example Edmondson and Cha (2006) argue that because mission statements are often non-specific and open to interpretation, workers will tend to expect more than employers intended to convey. If these stronger expectations take the form of expecting a stronger signal from better firms (coupled with the threat of believing them to be the worst type should they fail to send such a signal), then this could in fact increase the amount of information conveyed by mission statements. This may even be *ex ante* optimal for the principal, but nonetheless when the signaling game occurs the principal would always prefer a less costly equilibrium.

## 2.5 Example

The paper now turns to an example with specific functional forms in order to illustrate the equilibria in this game, and to consider the welfare implications of achieving more effective communication at the start of the game, at the potential cost of having to send wasteful signals at the start of the second period.

Assume that $\theta \sim U[0,1]$. Let the worker have utility function $U(a_1, a_2, \theta) = -(a_1 - \theta)^2 - (\frac{a_1}{2} + \frac{a_2}{2} - \theta)^2$, and let the principal have the profit function $\Pi = -(a_1 - b - \theta)^2 - (a_2 - b - \theta)^2$. These payoffs are similar to those in Crawford and Sobel (1984), in that each party wants the agent's actions to be matched to a function of the principal's type, $\theta$, but a bias $b$ is introduced that means the principal always prefers higher actions to the agent. These payoffs are extended to incorporate the agent's actions across both periods: whereas the principal wants the agent to exert effort $\theta + b$ in both periods, the agent will take into account his earlier effort $a_1$ when choosing second period effort $a_2$.

There are a couple of ways of thinking about these payoffs. For example, a behavioural argument could be made that the agent feels a sense of obligation to supply effort to the principal, and the extent of that obligation depends on $\theta$. Moreover if the agent supplied more effort to the principal in the earlier stages of the game, he will feel less of an obligation in later periods. We could think of the agent maximising a utility function of the form $U_t(a_1, ..., a_t, \theta) = -(\frac{1}{t}\sum_{s=1}^{t} a_s - \theta)^2$ in each period $t = 1, 2$ (with no discounting). That is, the agent wants to ensure in each period that his mean level of effort over the history of the game matches the obligation he feels to the principal. Alternatively, we could represent these payoffs in a per-period form $U_t(a_1, ..., a_t, \theta) = 2\theta(\frac{1}{t}\sum_{s=1} ta_s) - (\frac{1}{t}\sum_{s=1} ta_s))^2 - \theta^2$. We can think of this per-period payoff function as representing a salary which is a function of the principal's type and the agent's entire history of effort supplied, minus a cost of effort. This effort cost $c(a_1, ..., a_t) = -(\frac{1}{t}\sum_{s=1} ta_s))^2$ also depends on the history of effort provided: this dependence could come from the fact that having supplied higher effort in the past the worker is now less interested in his task, or because the opportunity cost of work has increased as the worker has spent less time on other activities. In this example the principal does not share these considerations, and so prefers the agent to supply higher effort in the second period, regardless of his first period action.

Given these payoffs, the agent would ideally set $a_1 = \theta$ in the first period. However the agent is uncertain about $\theta$, and given the principal's bias $b$ we cannot have fully informative communication in this setting. However there may exist a non-babbling cheap-talk equilibrium, in which case the agent's belief about $\theta$ will be informed by $m$. For this functional form the agent cares only about the expectation of $\theta$, not the distribution as a whole, and so $a_1^* = E[\theta|m]$. In the second period, the agent wishes to set the mean of his actions equal to $\theta$, and so $a_2^* = 2E[\theta|\mathscr{I}] - a_1^*$, where $\mathscr{I}$ indicates the agent's information in the second period. This information may comprise only the message sent at the start of the game, or it may also include the signal, if available. If the signal is unavailable, then the agent has no reason to update his beliefs at the start of the second period, and so $E[\theta|m, \varnothing] = E[\theta|m] \Rightarrow a_2^* = a_1^*$. If the signal is sent, then the agent learns $\theta$ with certainty in the fully-separating equilibrium, and so $a_2^* = 2\theta - E[\theta|m]$.

Turning to the principal's problem: if the signal is not sent, then the principal receives the same payoff in each period. Otherwise, with probability $p$, the principal must distinguish her type by sending a costly signal, following which the agent learns her type with certainty. Therefore the principal's payoff becomes $\Pi(\theta, m) = -(2-p)(E[\theta|m] - b - \theta)^2 - p(\theta - E[\theta|m] - b)^2 - ps(\theta, m)$. We will allow for both the least-cost fully separating equilibrium in the signaling game, and for more costly equilibria in which the indifferent type $\hat{\theta}_i$ sends a signal $\hat{s}_i$. If the signaling game occurs then the signal sent by a type $\theta \in [\hat{\theta}_{i-1}, \hat{\theta}_i]$ that sent a message $m_i$ must satisfy:

$$\theta = \arg\max_{\tilde{\theta}} \left\{ (2\tilde{\theta} - a_1^*(m_i) - b - \theta)^2 - s(\tilde{\theta}) \right\}$$

which implies that

$$s(\theta, m_i) = 4(\theta - \hat{\theta}_{i-1}) \left( \frac{\hat{\theta}_i + \hat{\theta}_{i-1}}{2} + b - \frac{\theta - \hat{\theta}_{i-1}}{2} \right).$$

This model nests the Crawford-Sobel payoffs, and so for $p = 0$ we know that no non-babbling cheap-talk equilibrium exists for $b > 1/4$. For non-zero values of $p$, however, the fact that the principal may have to follow up on the message with a costly signal, which will depend on the message sent. Consider the cheap-talk equilibrium with $n$ partitions $\{0, \hat{\theta}_1, ..., \hat{\theta}_i, ..., \hat{\theta}_{n-1}, 1\}$; we know from Proposition 1 above that the cheap-talk equilibrium will take this form. Let types in the interval $[\hat{\theta}_{i-1}, \hat{\theta}_i]$ send a message $m_i$. Upon observing this message the agent will take the first-period action $a_1^*(m_i) = E[\theta|m_i] = \frac{\hat{\theta}_{i-1}+\hat{\theta}_i}{2}$.

Given the cost that a type $\theta$ must bear if the signaling game occurs, we can solve for the indifferent types $\{0, \hat{\theta}_1, ..., \hat{\theta}_i, ..., \hat{\theta}_{n-1}, 1\}$

$$-(2-p)\left[\frac{\hat{\theta}_{i-1}+\hat{\theta}_i}{2} - \hat{\theta}_i - b\right]^2 - p\left[\hat{\theta}_i - \frac{\hat{\theta}_{i-1}+\hat{\theta}_i}{2} - b\right]^2 - p[\hat{s}_i + 4(\hat{\theta}_{i-1}+\hat{\theta}_i)(\hat{\theta}_{i-1}+b)] =$$

$$= -(2-p)\left[\frac{\hat{\theta}_i+\hat{\theta}_{i+1}}{2} - \hat{\theta}_i - b\right]^2 - p\left[\hat{\theta}_i - \frac{\hat{\theta}_i+\hat{\theta}_{i+1}}{2} - b\right]^2 - p[\hat{s}_{i+1}]$$

for $i = 1, ..., n-1$ where $\hat{\theta}_0 = 0$ and $\hat{\theta}_n = 1$.

For the simplest non-babbling cheap-talk equilibrium, that with 2 messages, there is an indifferent type $\hat{\theta} = \frac{1}{2} - \frac{2(b-\hat{s}p)}{4bp+1}$. As discussed above, the cost of signaling $\hat{s}$ is restricted such that $\hat{s} \in [0, 2\hat{\theta}(1/2+b)]$: the type $\hat{\theta}$ cannot be induced to send a signal greater than this, since she always has the option of sending a signal $s = 0$ and being believed to be the worst type. To confirm that this indeed constitutes an equilibrium, it must be verified that $\hat{\theta} \in (0,1)$, this implies that $b \leq \frac{1}{4(1-p)}$. For $p = 0$, therefore, an informative cheap-talk equilibrium exists only if $b < 1/4$, as in Crawford-Sobel 1994; however for any $p > 0$, there exist values of $b$ such that cheap-talk cannot be informative in the game without the opportunity to signal, but cheap-talk can be informative in the augmented game.

It is also worth noting that the result that $\hat{\theta} < 1/2$ holds here, even if we set the maximum signaling cost $s_{max}(\hat{\theta}, m_L) = 4\theta(1+2b)$. In this case $\hat{\theta} = \frac{1}{2} - 2b(1-p) < \frac{1}{2}$ for all $p > 0$. Therefore weaker messages convey more information, in the sense that the measure of the interval $[\hat{\theta}_{i-1}, \hat{\theta}_i]$ is decreasing in $i$.

Without stating the exact cutoffs here, for smaller $b$, larger $p$, or larger $\hat{s}$ we may also have more informative equilibria. Figures (1) and (2) below show the signals sent in such equilibria, with 2 and 3 messages respectively. The images include both the lowest cost equilibrium, in which the weakest type in each interval sends a costless signal, and the most costly equilibrium, in which the lowest type in an interval sends a signal equal to that of the highest type in the adjacent interval. As discussed in section 4.2, increasing the cost of the signal the weakest type in each interval sends makes imitating a higher type less appealing for weaker firms, shifting the cutoffs upwards and increasing the informativeness of communication.

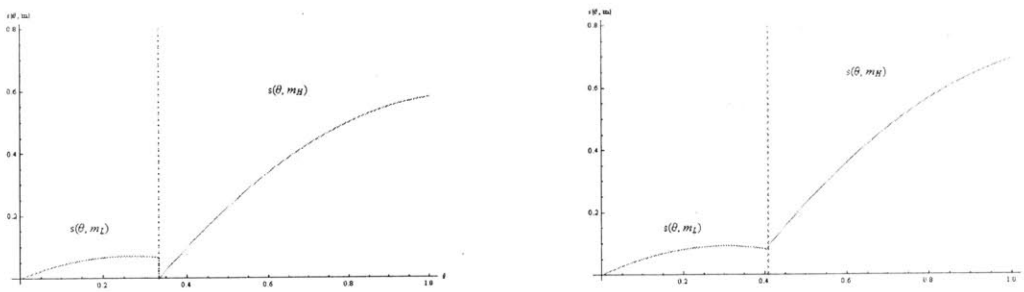


Figure 2-1: Signals sent in equilibrium with two messages: Costless (L) and Costly (H)

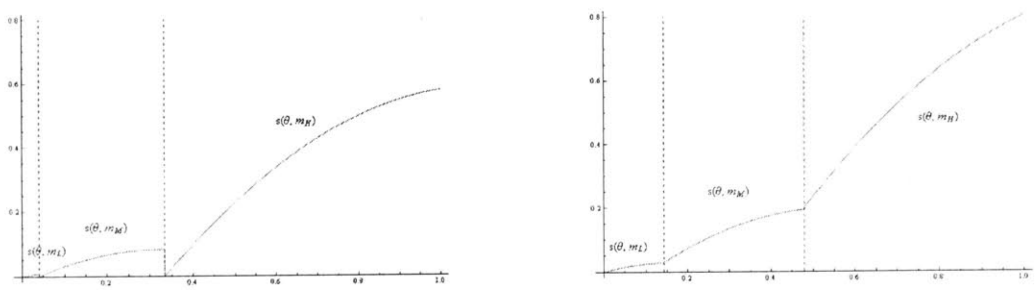


Figure 2-2: Signals sent in equilibrium with three messages: Costless (L) and Costly (R)

In both the equilibrium with two messages and that with three, increasing the cost of signals sent by those types in the higher interval increases the informativeness of the equilibrium, in the sense that the variance of the agent's belief decreases. Whilst this represents an increase in the agent's welfare, the effect on the principal's total profits, net of the signal, is ambiguous. The paper turns to the welfare implications in the next section.

### 2.5.1 Welfare

This section considers the welfare implications of inducing more informative communication. As discussed, increasing either the probability of having to send a costly signal $p$ or selecting an equilibrium in which types in higher intervals must send more costly signals, may generate an equilibrium in which more messages are sent, and thus more information conveyed to the agent. We can therefore ask whether taking measures to increase these expected signaling costs can increase total welfare. For example, we may imagine that increasing the scope of these messages or sending them to a wider audience may increase the probability that a firm is called upon to act. This accords with the suggestion in the management literature that messages that are broader, rather than focused on a particular aspect of the firm such as sales or technology, are more effective. The cost of this, however, is that since the signal represents money-burning, it is wasteful in terms of total surplus.

In this section I show that increasing the number of messages sent improves expected utility for the agent and, discounting the signal, also increases *ex ante* expected welfare for the principal. However once the signaling costs are accounted for the total effect on the principal's welfare, and on total welfare, may be ambiguous. This result is in accordance with some of the arguments made in the management literature regarding the drawbacks of mission statements. It is argued that workers may interpret mission statements in a particular way, and be quick to attribute failures to live up to these interpretations to hypocrisy. This disappointment in the firm is something we would expect to see in equilibrium: in any interval there will be types that fall below the mean of those sending a particular message, and the agent will react to this by reducing the effort sent. Moreover, to the extent that the agent can influence the signaling equilibrium through his beliefs and induce an equilibrium which is more informative, but more costly to the principal, it may be beneficial to the agent to interpret the messages in a way that places more demands on the principal.

**Proposition 6.** *The agent's utility $U(m,\theta)$ and the principal's profits $\Pi(a_1,a_2,\theta)$ (non-inclusive of signaling costs) are increasing in the number of messages sent in equilibrium.*

*Proof.* From Proposition 4 above, we know that increasing the number of messages sent from $n$ to $n' > n$ means that each type $\theta$ will send a message that reveals the firm to be in an interval $[\hat{\theta}'_{i-1}, \hat{\theta}'_i]$ of smaller measure than the interval $[\hat{\theta}_{i-1}, \hat{\theta}_i]$. I now show that decreasing the size of the interval increases the agent's expected utility, conditional on the message sent.

$$\begin{aligned} E[U(a_1,a_2,\theta)|m_i] &= -(2-p)\int_{\hat{\theta}_{i-1}}^{\hat{\theta}_i}\left(\frac{\hat{\theta}_{i-1}+\hat{\theta}_i}{2}-\theta\right)^2 d\theta \\ &= -\frac{2(2-p)}{3}\left(\frac{\hat{\theta}_i+\hat{\theta}_{i-1}}{2}\right)^2. \end{aligned} \tag{2.9}$$

Since the agent's expected utility increases as the size of each interval decreases, and all types will send a message corresponding to a smaller set of types, the agents ex ante expected utility is increasing in $n$.

Similarly, turning to the principal's profits, non-inclusive of the signaling costs

$$\begin{aligned} E\Pi &= -(2-p)\int_{\hat{\theta}_{i-1}}^{\hat{\theta}_i}\left(\frac{\hat{\theta}_{i-1}+\hat{\theta}_i}{2}\right)^2 d\theta - p\int_{\hat{\theta}_{i-1}}^{\hat{\theta}_i}\left(\theta-\frac{\hat{\theta}_{i-1}+\hat{\theta}_i}{2}-b\right)^2 d\theta \\ &= -\frac{4}{3}(\hat{\theta}_i-\hat{\theta}_{i-1})((\hat{\theta}_i-\hat{\theta}_{i-1})^2+3b^2). \end{aligned} \tag{2.10}$$

The principal's profits are therefore also increasing in $\hat{\theta}_i - \hat{\theta}_{i-1}$, and so increasing the number of messages sent increases expected profits to the principal. □

Therefore the agent is always better off by endeavouring to select a more informative equilibrium. This result will extend to any risk-averse payoff function. To the extent that the agent can influence equilibrium selection, for example through the beliefs that they will ascribe to different signals, the agent thus has an incentive to try and induce the principal to take more costly actions if the costly signal is available.

As far as increased information is concerned, the principal's and the agent's beliefs are aligned; if we ignore the signaling costs, then the principal's expected profits ex ante are also increasing in the number of messages sent. That is, before knowing her type, the principal would have a preference for revealing as much information as possible through the cheap-talk messages. However once her type is drawn, the bias $b$ means that the principal always has a preference to distort communication, limiting the number of messages that may be sent in equilibrium. Once we account for signaling costs the net effect on both profits and total surplus is ambiguous. Increasing the number of partitions affects the signaling cost in two ways. Firstly, there is a direct effect in that it may be necessary to increase $\hat{s}$ in order to increase the number of messages sent in equilibrium. Secondly, for fixed $p$ and $\hat{s}$, selecting a more informative equilibrium increases the signaling costs by increasing the marginal cost of revealing their type to firms in higher intervals.

The figures (3), (4) and (5) show the net profits, expected utility and surplus from the non-babbling equilibrium, the equilibrium with 2 messages, and the equilibrium with 3 messages. Both the least-cost and most costly equilibria are shown. For the principal, increasing the number of partitions always increases the utility of the weakest type; this must be the case, since moving from an equilibrium in which $n$ messages may be sent to one in which $n+1$ messages may be send requires that the weakest type (in this case $\theta = 0$) is willing to announce his true type. However in this example the strongest types are made worse off in more informative equilibria: whilst the increase in information increases their payoffs from the agent's actions, this comes with the downside of a more costly signal. The firms who make the strongest commitments through their mission statements induce the highest expectations in their employees, and are therefore constrained to bear greater costs to maintain this reputation.

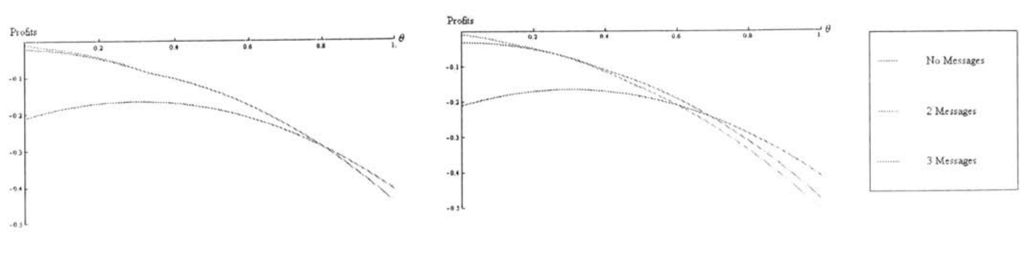


Figure 2-3: Profits in cheap-talk equilibrium: Costless (L) and Costly (R)

Although the net effect on the firm is unclear, improved communication will increase the expected utility of the employee. In particular, for the quadratic loss function in the example presented here, increasing the number of signals decreases the measure of each partition, and thus reduces the maximum loss that the agent may sustain. With more accurate information the agent is both able to set his first-period action closer to the full-information optimum, and in the event the signal is received, the variance of his second-period action will decrease. This is utility-improving for any risk-averse agent.

Figure (5) shows that the net effect is nonetheless ambiguous. Whilst moving from the babbling equilibrium to one in which the type-space is partitioned into two is welfare improving, refining the partitions decreases total surplus for most types $\theta$. In the case shown here, moving from two messages sent in equilibrium to three messages generates a slight increase in total surplus, for both the least costly and most costly equilibria, however this will not always be the case.

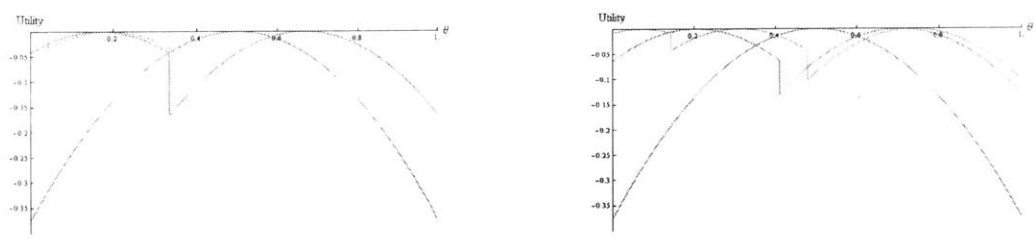


Figure 2-4: Agent's Utility by Firm Type: Costless (L) and Costly (R) Equilibria

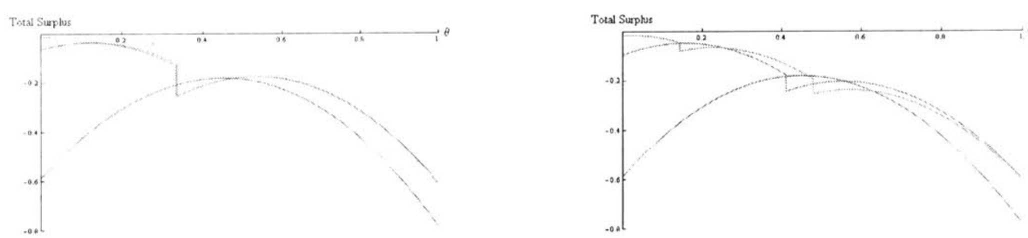


Figure 2-5: Total Surplus: Costless (L) and Costly (R) Equilibria

## 2.6 Adaptation

One of the drawbacks of mission statements discussed in the literature concerns problems of adaptation: circumstances may require a change in firm policy that could be interpreted as a betrayal of their mission statement. In this case workers may react angrily, even if the firm has been acting in good faith and did not anticipate a change in direction. If the firm is aware that they might need to change their policies they will be reluctant to make any specific promises in their mission statements and thus we will expect communication to be more generic and less informative.

This situation will be modeled by assuming that with probability $\gamma$ the principal's type changes, and is redrawn independently from $U[\underline{\theta}, \bar{\theta}]$. I will assume that the agent does not know whether a new draw has taken place. For simplicity of exposition, I will focus on the specific example from section (5), however these results will hold for the more general case.

In the first period, the agent will choose effort as before and so $a_1^*(m) = E[\theta|m]$. In the second period, if the signal is sent then the agent learns the principal's type (but not whether the type has changed) with certainty, and in that case will set $a_2^*(m, s) = 2\theta - E[\theta|m]$. If the signal is not sent, then the agent receives no new information about the principal's type, but must take into account the possibility that the principal's type has changed and so the original message is uninformative. Note that if the principal's type changes and the signal is unavailable there is no scope for further information to be conveyed through cheap-talk messages: there is no possibility of another signal being required in the future and so there would only be a babbling equilibrium in this game. Therefore the agent will update his belief about $\theta$ to $E[\theta|m, \varnothing] = \gamma E[\theta] + (1-\gamma)E[\theta|m]$, and chooses second-period action $a_2^*(m, \varnothing) = \gamma + (1-2\gamma)E[\theta|m]$.

For the principal, the possibility of a change in type not only changes the value of the initial message sent, but also affects the signaling costs. This is because, regardless of the message sent, in the second period the principal must now distinguish herself from all potential types, not only those who sent messages lying in a particular interval. This means that the relevant "lowest type" is now $\underline{\theta} = 0$ for all messages sent, and therefore the signal becomes $s(\theta, m) = 2\theta\left(a_1 + b - \frac{\theta}{2}\right)$.

I now consider the cheap-talk game, and the principal's expected payoff from sending a message $m_i$, given her original type $\theta$. The principal's payoff from the agent's first-period action will be unchanged. In the second period, if no signal is available the agent will choose his action based on the message and taking into account the possibility that the principal's type will change. The principal must also anticipate that with probability $\gamma$ her type will change. If the signal is available, the agent will learn the principal's type with certainty, whether or not it is redrawn but the principal's expected profit and signal must take into account this possibility. The principal's expected profit is therefore:

$$\begin{aligned}\Pi(\theta, m) = &- (a_1(m) - b - \theta)^2 - (1-p)\left[\gamma \int_0^1 (a_2^*(m, \varnothing) - b - x)^2 dx + (1-\gamma)(a_2^*(m, \varnothing) - b - \theta)^2\right] \\ &- p\left[\gamma \int_0^1 [(x - a_1(m) - b)^2 - s(x, m)]dx + (1-\gamma)[(\theta - a_1(m) - b)^2 - s(\theta, m)]\right].\end{aligned}$$

For a non-babbling cheap-talk equilibrium to exist, there must be at least one type $\hat{\theta} \in (0, 1)$ that is indifferent between sending a lower message $m_L$, inducing the belief that $\theta \sim U[0, \hat{\theta})$, and the higher message, inducing belief $\theta \sim U[\hat{\theta}, 1]$. Solving for the indifferent type in this example with two messages, $\hat{\theta} = \frac{2-\gamma(4pb-1)}{4(1-p)(1-2\gamma^2)} - \frac{\gamma^2+2b(1-\gamma)}{1-2\gamma^2}$. This is decreasing in $\gamma$, so as the probability that the principal's type changes increases, the informativeness of communication decreases.

For more general payoff functions, increasing $\gamma$ makes it more tempting for a low type to attempt to imitate a higher type by sending a higher message: given that there is some probability that the firm's type will change, and in expectation increase, the firm is less likely to bear the consequences of imitating a higher type and then revealing themselves to be of lower quality through the signal. For higher types, given the possibility that their type will change and they will be unable to live up to promises made in good faith at the start of the game, the payoffs from attempting to distinguish themselves will decrease. Even if the signal is not available, the agent will take into account the possibility that the principal's type has been redrawn and thus, for any message sent at the start of the game, his beliefs will revert towards the mean in the second period. This also reduces the gains for the principal of trying to convey information about her type.

## 2.7 Conclusion

This paper presents a model of cheap-talk in which communication may be informative if with positive probability the sender has the opportunity to follow through on her promises through a costly signaling mechanism. I argue that this may be interpreted as a model of mission statements by firms, whereby non-binding, and seemingly platitudinous, promises and commitments can carry weight if they constrain the firm's future actions and induce the firm to make greater sacrifices in order to convince employees of their quality.

I focus on cheap-talk games in which, absent the possibility of sending a costly signal later in the game, a non-babbling equilibrium would not exist. I show that if the probability of the firm being able to distinguish its type through a costly signal is sufficiently high, initial cheap-talk messages may be able to convey some information by partitioning the type-space into distinct intervals. Moreover, further increases in $p$ will increase the informativeness of such a cheap-talk equilibrium both by weakly increasing the number of messages that may be sent in equilibrium, and by reducing the expected variance of the agent's beliefs conditional on these messages.

Although increasing the probability that the firm has an opportunity to signal will increase the amount of information conveyed to the agent, this paper shows that the overall welfare effect may be ambiguous. Since the signal may be interpreted as money burning, it is wasteful from a social point of view, and these losses may offset the increase in expected utility and profits due to better information. There are further costs associated with these cheap-talk messages if the possibility that the firm's type or mission will change in response to changing circumstances: not only does this reduce the information conveyed by messages at the start of the game, but it increases the costs that the firm must bear to reveal their type should the signal become available, and may increase the "punishment" on the firm for a perceived betrayal.

## 2.8 Appendix

### 2.8.1 Proof of Proposition 1

*Proposition: A non-babbling cheap-talk equilibrium that partitions the type-space into two intervals will exist in this game if there exists a type $\hat{\theta}$ that is indifferent between sending the lower message and being believed to be in the set $[\underline{\theta}, \hat{\theta}]$, and sending the higher message and being believed to be in the set $[\hat{\theta}, \bar{\theta}]$. The type $\hat{\theta}$ must satisfy:* $(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] = (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s(\hat{\theta}), \hat{\theta})]$.

*Proof.* For now, consider the limited message space $\mathscr{M} = \{m_L, m_H\}$. If pure-strategy cheap-talk is informative, then it will separate the type-space into two subsets $\Theta_L$ and $\Theta_H$, where $\Theta_L \cup \Theta_H = \Theta$ and $\Theta_L \cap \Theta_H = \varnothing$. In equilibrium, messsages are sent according to a function $\phi : \Theta \to \mathscr{M}$, such that types $\theta \in \Theta_i$ send messages $m_i$. That is, types in the disjoint sets $\Theta_L$ and $\Theta_H$ send distinct messages. Let $f(\theta|m_i)$ be the agent's belief on observing a message in $m_i$.

I now show that the partition will be into two intervals $[\underline{\theta}, \hat{\theta})$ and $[\hat{\theta}, \bar{\theta}]$. Assume that $f(\theta|m_i)$, for $i = L, H$, is such that $a_1^*(m_H) > a_1^*(m_L)$. Consider some $\theta$ such that

$$\begin{aligned}(1-p)\Pi(a_1(m_H), a_2(a_1(m_H), m_H), \theta) + p[\Pi(a_1(m_H), a_2(a_1(m_H), s_H(\theta)), \theta) - s(m_H, \theta)] > \\ (1-p)\Pi(a_1(m_L), a_2(a_1(m_L), m_L), \theta) + p[\Pi(a_1(m_L), a_2(a_1(m_L), s_L(\theta)), \theta) - s(m_L, \theta)],\end{aligned}$$

then for any $\theta' > \theta$, this expression also holds. Taking the first part of each expression, since the profit function satisfies increasing differences in $a_1$ and $\theta$, $\Pi(a_1(m_H), a_2(a_1(m_H), m_H), \theta') - \Pi(a_1(m_H), a_2(a_1(m_H), m_H), \theta) > \Pi(a_1(m_L), a_2(a_1(m_L), m_L), \theta') - \Pi(a_1(m_L), a_2(a_1(m_L), m_L), \theta)$. For the remainder of the expression, we require

$$\Pi(a_1(m_H), a_2(a_1(m_H), s(\theta)), \theta) - s(m_H, \theta) - (\Pi(a_1(m_L), a_2(a_1(m_L), s(\theta)), \theta) - s(m_L, \theta))$$

to be increasing in $\theta$. That is, should the signaling game occur, the gain from having sent the higher message is increasing in $\theta$, where

$$\frac{d}{d\theta}[\Pi(a_1(m), a_2(a_1(m), s(\theta)), \theta) - s(m, \theta)] = \frac{\partial \Pi}{\partial a_2}\frac{\partial a_2}{\partial \theta} + \frac{\partial \Pi}{\partial \theta} - \frac{ds(m,\theta)}{d\theta} = \frac{\partial \Pi}{\partial \theta}.$$

We know that $\frac{\partial^2 \Pi}{\partial a_1 \partial \theta} > 0$, and so the gains from having sent a higher message and induced a higher action $a_1$ are increasing in $\theta$, as desired.

The converse is also true: that if some type $\theta$ prefers to send the lower message, then any type $\theta' < \theta$ will also prefer to send the lower message. Thus we know that if there exists some $\hat{\theta}$ for which this expression holds with equality, i.e., that is indifferent between sending the lower message $m_L$ and the higher message $m_H$, then all types $\theta > \hat{\theta}$ will strictly prefer to send the higher message, and all types $\theta < \hat{\theta}$ will strictly prefer to send the lower message.

Therefore messages will partition the type space into two subsets $[\underline{\theta}, \hat{\theta}]$, $[\hat{\theta}, \bar{\theta}]$. Moreover, since the messages will partition $\Theta$ into two disjoint intervals we know that these messages and the concomitant beliefs will in fact induce effort $a_1^*(m_1) > a_1^*(m_2)$.

We can extend this proof to allow for a more general message space, $\mathscr{M}$, by assuming that types in the sets $\Theta_1$ and $\Theta_2$ send messages drawn from the appropriate subset $\mathscr{M}_i \subset \mathscr{M}$, where $\mathscr{M}_1 \cap \mathscr{M}_2 = \varnothing$. □

### 2.8.2 Proof of Lemma 1

*Lemma: If* $(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] < (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s(\hat{\theta}), \hat{\theta})]$ *for some* $\hat{\theta}$*, then this is also true for all* $\hat{\theta}' > \hat{\theta}$.

*Proof.* The statement is equivalent to

$$p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta})] - ps(m_L, \hat{\theta}) < (1-p)[\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta})$$

for some $\hat{\theta}$. Consider some $\hat{\theta}' > \hat{\theta}$. Increasing the cutoff point between the two messages increases the agent's expectation of $\theta$, conditional on either message. Moreover, we must have $a_1^*(m_H, \hat{\theta}') - a_1^*(m_L, \hat{\theta}') > a_1^*(m_H, \hat{\theta}') - a_1^*(m_L, \hat{\theta}')$. This comes from the fact that there are increasing differences in $a_1$ and $\theta$ in the agent's payoff function. This is turn implies that $\Pi(m_H, \hat{\theta}') - \Pi(m_L, \hat{\theta}') > \Pi(m_H, \hat{\theta}') - \Pi(m_L, \hat{\theta}')$. Therefore the right-hand side of the expression above must be increasing in $\hat{\theta}$. Turning to the left-hand side of the expression, we know that $s(m_L, \hat{\theta})$ must be increasing in $\hat{\theta}'$. This comes not only from the fact that the higher type $\hat{\theta}'$ must pay more to distinguish itself from the type $\hat{\theta}$, but moreover from the fact that increasing $\hat{\theta}$ also increases $a_1^*(m_L)$. This effectively increases the stakes for the principal, and raises the marginal cost of signaling. Moreover we must have $\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta})$ decreasing in $\hat{\theta}$, since increasing differences between $a_1$ and $\theta$ implies that $\Pi(m_H, s(\hat{\theta}'), \hat{\theta}') - \Pi(m_H, s(\hat{\theta}), \hat{\theta}) > \Pi(m_L, s(\hat{\theta}'), \hat{\theta}') - \Pi(m_L, s(\hat{\theta}), \hat{\theta})$. Therefore equation 2 will still hold if the cutoff is increased from $\hat{\theta}$ to $\hat{\theta}'$. □

### 2.8.3 Proof of Lemma 2

*Lemma: If a non-babbling cheap-talk equilibrium exists in which the type-space is partitioned into two, then it must be the case that the indifferent type* $\hat{\theta} < \frac{\underline{\theta}+\bar{\theta}}{2}$.

*Proof.* For the indifferent type $\hat{\theta}$ it must be the case that

$$(1-p)\Pi(m_L, \hat{\theta}) + p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta})] = (1-p)\Pi(m_H, \hat{\theta}) + p[\Pi(m_H, s(\hat{\theta}), \hat{\theta})]$$

which implies that

$$p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta})] - ps(m_L, \hat{\theta}) = (1-p)[\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta}). \quad (2.11)$$

We know that if the signaling game does not occur then the principal cannot be punished for sending a misleading message, and so the right-hand side of this expression is strictly positive for any $p > 0$. We want to show that if $\hat{\theta} > \frac{\underline{\theta}+\bar{\theta}}{2}$ then

$$p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta})] - ps(m_L, \hat{\theta}) < (1-p)[\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta})$$

for any $p$. To consider the most extreme case, consider $p = 1$: if this expression is satisfied for $p = 1$, it must hold for all $p$. In particular, we want to show that

$$\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - \Pi(m_H, s(\hat{\theta}), \hat{\theta}) < 0 \text{ if } \hat{\theta} \geq \frac{\underline{\theta}+\bar{\theta}}{2}. \tag{2.12}$$

From Assumption 1, we have

$$\frac{d\Pi(a_1(\tilde{\theta}), a_2(a_1(\tilde{\theta}), \theta), \theta)}{d\tilde{\theta}} > 0$$

for all $\tilde{\theta} < \theta$. It follows that

$$\Pi(a_1(m_L), a_2(a_1(m_L), \hat{\theta}), \hat{\theta}) < \Pi(a_1(\hat{\theta}), a_2(a_1(\hat{\theta}), \hat{\theta}, \hat{\theta})$$

since $m_L$ must induce a belief lower than the principal's true type $\hat{\theta}$. We want to compare this loss in surplus from imitating a lower type to the change in surplus from imitating a higher type (given that the payoffs from imitating a type $\tilde{\theta}$ are still increasing at $\tilde{\theta} = \theta$). Even if the principal is made worse off by sending the higher message, if $\hat{\theta} - \underline{\theta} > \bar{\theta} - \hat{\theta}$ then $\hat{\theta} - E[\theta|m_L] > E[\theta|m_H] - \hat{\theta}$ and from the concavity of the principal's profit function in $\tilde{\theta}$ (assumption 3) it must be the case that:

$$\left|\Pi(a_1(m_L), a_2(a_1(m_L), \hat{\theta}), \hat{\theta}) - \Pi(a_1(\hat{\theta}), a_2(a_1(\hat{\theta}), \hat{\theta})\right| >$$
$$\left|\Pi(a_1(m_H), a_2(a_1(m_H), \hat{\theta}), \hat{\theta}) - \Pi(a_1(\hat{\theta}), a_2(a_1(\hat{\theta}), \hat{\theta})\right|$$

given the beliefs induced. Therefore even if the principal's payoffs after the signaling game from sending the higher message are less than those from announcing her true type (given that the agent punishes the deception by withdrawing effort), she must still be better off than if she imitates the low type. This will be true for any $\hat{\theta} > \frac{\underline{\theta}+\bar{\theta}}{2}$. So expression (2.12) must hold, and therefore equation (2.11) cannot be satisfied for any $\hat{\theta} \geq \frac{\underline{\theta}+\bar{\theta}}{2}$, even for $p = 1$ and ignoring the additional costs of signaling from sending a lower message. For $\hat{\theta} < \frac{\underline{\theta}+\bar{\theta}}{2}$ the true type will be closer to the agent's expectation conditional on the message $m_L$, and so this argument from concavity will no longer hold.

*This result will still hold for higher signaling costs*

If we consider the equilibrium in which the weaker type in the higher interval must send a costly signal, it will still be the case the that indifferent type $\hat{\theta}(s) < \frac{\underline{\theta}+\bar{\theta}}{2}$. In order for this to be the case, we require:

$$p[\Pi(m_L, s(\hat{\theta}), \hat{\theta}) - s(m_L, \hat{\theta}) - (\Pi(m_H, s(\hat{\theta}), \hat{\theta}) - s(m_H, \hat{\theta}))] = (1-p)[\Pi(m_H, \hat{\theta}) - \Pi(m_L, \hat{\theta}). \tag{2.13}$$

As above, consider $p = 1$, the case in which the signaling game occurs with certainty. In this case the signal $s(m_H, \hat{\theta})$ is defined as

$$s(\hat{\theta}, m_H) = s(\underline{\theta}) + \int_{\underline{\theta}}^{\hat{\theta}} \frac{d}{d\tilde{\theta}} \Pi(a_1(m_H), a_2(a_1(m_H), s(\tilde{\theta}, m_H)), x)\Big|_{\tilde{\theta}=x} dx \geq s_{max}(\hat{\theta}, m_H).$$

We want to show that the left-hand side of expression (2.13) must be negative for $\hat{\theta} \geq \frac{\underline{\theta}+\bar{\theta}}{2}$. We know that

$$\frac{d^2}{d\theta^2}[\Pi(a_1(m), a_2(a_1(m), s(\theta)), \theta) - s(m, \theta)] = \frac{d}{d\theta}\left[\frac{\partial \Pi}{\partial a_2}\frac{\partial a_2}{\partial \theta} + \frac{\partial \Pi}{\partial \theta} - \frac{ds(m,\theta)}{d\theta}\right] = \frac{d}{d\theta}\frac{\partial \Pi}{\partial \theta} < 0$$

and so the expected payoffs after the signaling game, inclusive of signaling costs, are concave in $\theta$. We also know from assumption 1 that the payoffs to imitating a type $\tilde{\theta}$ are increasing in $\tilde{\theta}$ for all $\tilde{\theta} < \theta$, and so a principal of type $\theta$ must strictly prefer to reveal his true type than imitate a weaker type. From the concavity of his payoffs in $\theta$, it therefore follows that if $\hat{\theta} - E[\theta|m_L] > E[\theta|m_H] - \hat{\theta}$ he must be strictly worse off sending the lower message than the higher message, even if the signaling game occurs with certainty. Therefore we must have $\hat{\theta} < \frac{\underline{\theta}+\bar{\theta}}{2}$.

□

# Chapter 3

# Matching the Market: Relational Contracts and the Dynamics of Bonus Payments

## 3.1 Introduction

The payment of large bonuses, especially in the finance industry, has been a subject that has attracted debate and opprobrium over recent years. Whilst most would acknowledge the value of making these payments as a way of rewarding performance, they are often seen as being disproportionately generous. These criticisms have become more salient as bonus payments have rapidly rebounded following a period of economic turmoil, especially since many industries remain fragile. However bonus payments are not only a means of providing incentives, they may also serve to convince workers of the quality of their employer, and will be subject to competitive and market pressures.

This paper presents a model in which bonus payments not only provide incentives for effort but may also reveal information about the principal's type. I will assume that a principal of privately known type seeks to elicit effort from an agent in an infinitely repeated game. Both the principal and the agent will be risk-neutral and will share a common discount factor. However the principal may be one of two types, where her type determines the probability that high output is produced, conditional on the agent supplying effort. I will assume that output is non-contractible but is observed by both parties and so high effort can be supported by a relational contract in which a bonus is paid following high output. In this setting the agent will update his beliefs about the principal's type based on his observations of output and on the bonus received.

This paper shows that since the agent must update his beliefs about the principal's type based on his observations of output the two types cannot pool indefinitely. I focus on the case in which the low type is unable to obtain effort from the agent when her type is known, and show that in this case the two types cannot separate through the contract offered, but that the high type may be able to separate by paying a larger bonus following the first instance of high output. Alternatively, the two types may initially pool and both pay the same bonus, waiting for the agent to update his beliefs over time. In both cases the necessary bonus payment will evolve with the agent's belief, and there is positive probability that following a run of low output the agent's belief may fall to a level at which he is no longer willing to supply effort. This implies that even a high-type principal may find herself trapped in an inefficient equilibrium. The desire to avoid this outcome may lead the principal to pay bonuses greater than those required to elicit effort in order to induce separation at the earliest opportunity, especially after a run of low output.

One way of thinking about this setting is as a model of firm resilience. We can imagine that the output of the firm depends on the agent's effort, some external shock, and the ability of the firm to resist or adapt to this shock. A better firm is therefore able to withstand a greater proportion of shocks, and will therefore generate high output more frequently, despite external events. Firms will try to signal their type through their ability to pay larger bonus payments. In particular we will expect to see higher bonus payments after a run of bad output as the firm must restore the agent's belief in its quality. A run of low output will also increase the pressure on good firms to reveal their type, since their relationship with the agent will become more fragile as his beliefs deteriorate. This model therefore predicts that bonuses may dramatically increase at the earliest opportunity following a run of poor performance.

The paper goes on to more explicitly model how the dynamics of bonus payments may be affected by the agent's perceptions of markets conditions. I augment the model to include a signal, observed by both parties, that indicates whether a high type principal would be expected to produce high output in a particular period. This signal may be interpreted as an indicator of market conditions, or the performance of potential competitors. The agent's beliefs about the principal's type will therefore also depend on the signal received. In particular, if the signal suggests that other firms are struggling then the agent will be more lenient in his interpretation of low output, whereas if others appear to be performing well the agent will interpret low output more harshly and may either cease to supply effort or demand much larger bonus payments in the future to restore his confidence. In addition to affecting the agent's beliefs and hence the bonus that must be promised in future periods, the signal may also influence the principal's decision to pay a bonus in the current period. We will expect to see scenarios in which the principal pays the bonus following a positive signal in an attempt to match the market and keep up with her competitors, whereas she may have reneged had the signal been unfavourable.

As a possible example of some of these dynamics, we may consider the case of Credit Suisse First Boston, reported in the New Yorker (Stewart, 1993). Widely regarded as a leading investment bank, First Boston were taken private by Credit Suisse in 1987, causing some concern over whether performance and compensation would continue to match the levels of other Wall Street banks. Although performance was poor in the late 1980s, this initially was not an issue: other firms were also struggling and so employees were prepared to accept lower levels of bonuses. However when bonus payments failed to match those of their perceived competitors as the economy recovered this occasioned widespread resentment as workers felt that they had been misled by the firm. These circumstances can be considered in the light of the model presented in this paper: during a run of poor performance, when the leading firms were also seen to be struggling, the firm faced little pressure to pay larger bonus payments, and workers were prepared to wait before evaluating the company. However once other firms were seen to be faring better Credit Suisse faced more pressure to come up with the generous bonus payments that would convince the workers to trust their quality. Within this setting, there are two ways of interpreting the firm's behaviour: it could be argued that Credit Suisse was in fact unable to match the levels of performance of their competitors, and so was unwilling to pay the larger bonuses which would allow them to continue to pool with these types. Alternatively, it may be that Credit Suisse acted in good faith: in this interpretation they would in fact be a high type, but the signals received by the workers from the market suggested that in this case they should have performed better. Therefore the failure to match their competitors in performance of bonus payment diminished the worker's trust in the firm to the point that they were no longer willing to supply effort.

This paper builds on the literature on relational contracts, such as Levin (2003); it also follows Halac (2011), in considering the extent to which a principal of privately known type is able to reveal her type through a relational contract. In that paper it is the principal's outside option that is unknown, whilst a pooling equilibrium will exist, the high type can also induce separation through paying a larger bonus. In this paper it is the probability that high effort generates high output that is privately known, and this implies that there must be some updating of the principal's beliefs even if the principal does not attempt to induce separation through a larger bonus. This in turn implies that no pooling equilibrium will exist in the long-run, and moreover that the types may fail to separate. The dynamics of the expected bonus payment seen in this paper can also be compared to those predicted in Li and Matouschek (2013). In both cases we would expect to see bonus payments increase following a run of poor output. However in their paper this is driven by temporary shocks to the principal's ability to pay, and the need to ensure that the principal reports truthfully. In this paper the principal's type is persistent, and affects her incentives, rather than her ability, to pay.

This paper also relates to the literature on relationship building. The results presented here suggest that high-type firms can convey their type to the worker, and hence gain the worker's trust, either through a run of success and gradual updating of the agent's beliefs, or failing that, by paying a much larger bonus. There is a substantial literature in which it is the agent's type that is unknown to the principal, and this will drive the dynamics of the relationship. For example Fong and Li (2010) consider a case in which a worker may go through an initial probationary period before the relationship is either terminated, or the worker is tenured. Watson (1999, 2002) considers the case in which the stakes in the relationship may gradually increase over time as more information is revealed.

The paper will proceed as follows. In the next section I outline the model. In Section 3 I analyse the possible equilibria in the baseline model, without signals, and consider the dynamics of the bonus payments made in equilibrium. In Section 4 I consider an augmented model in which the agent receives some signal of whether a high-type principal should be able to generate high output in a particular period, and I discuss how this changes the principal's incentives to pay the bonus. Section 5 concludes.

## 3.2 Model

A principal (she) and an agent (he), both of whom are risk neutral and who share a common discount factor $\delta$, interact in an infinitely repeated game. In each period, the agent chooses a level of effort $e \in \{0, 1\}$, where the cost of effort $c(e) = ce$. Output $y_t \in \{Y_L, Y_H\}$ is a function of the agent's choice of effort, the state of the world, and the principal's type $\theta$. Assume that the principal may be either a high or low type $\theta \in \{\theta_L, \theta_H\}$, where $\Pr[\theta = \theta_H] = p$. The principal's type will determine the probability that high output is generated, given the agent's choice of effort. Specifically, assume that $\Pr[y = Y_H | e = 0, \theta] = 0$ for all $\theta$, and $\Pr[y = Y_H | e = 1, \theta] = \pi_i$, $i = L, H$. Let $\pi_H > \pi_L$.

Thus if the agent supplies no effort, there is zero probability of high output being produced. If the agent supplies effort 1, at a cost $c$, then there is a positive probability of high output. This probability is greater for the high type of principal, and so the expected value of effort is higher. I will assume that $\pi_L(Y_H - Y_L) > c$, and so the expected value of output is such that it is worth supplying costly effort, regardless of the principal's type.

Assume that the principal cannot observe the level of effort supplied by the agent, but that both parties can observe output. However output will not be formally contractible. Effort in this game can therefore only be sustained by a relational contract and the promise of a bonus payment conditional on output. The compensation offered to the worker will take the form $w_t = s_t + b_t(y_t)$, where $s_t$ is a fixed salary, and $b_t$ a non-contractible bonus that may be contingent on output.

In each period, the timing of the game is as follows:

1. The principal offers the agent a contract specifying a salary $s_t$ and a bonus $b_y$, to be paid following high output. The agent accepts or rejects this offer. If the agent accepts then the salary $s_t$ is paid.

2. The agent chooses effort $e_t \in \{0, 1\}$, which is unobserved by the principal.

3. Output $y_t \in \{Y_H, Y_L\}$ is realised, and observed by both parties

4. The principal chooses whether or not to pay the bonus $b_t$.

5. Payoffs are received: $U = s_t - ce_t + b_t\mathbb{I}_{Bonus\ paid}$, $\Pi = y_t - s_t - b_y\mathbb{I}_{Bonus\ paid}$.

I will assume that the agent's outside option is $U_0 < Y_L$, and that the agent is drawn from an infinite set of identical workers, all of whom can observe the full history of output generated and transfers received. This assumption implies that the principal has the bargaining power in the relationship, and can hire a new agent should the relationship terminate. The principal's outside option is therefore bounded below by $Y_L$ in every period, the payoff she can obtain by hiring a new agent who supplies zero effort. I will assume that the agent has limited liability and therefore that any bonus payments must be non-negative.

## 3.3 Equilibrium

In this section I analyse the potential equilibria of the game, and whether, and how, the principal's type may be revealed. I show that in the long-run there cannot be a fully pooling equilibrium in which the agent supplies effort; this is because the agent must update his beliefs about the principal's type based on his observation of output. I then focus on the case in which, were the principal's type known, the low type would be unable to elicit effort from the agent. I show that in this case there cannot be a separating equilibrium which induces separation at the very beginning of the game, through the contract offered, but that there is an equilibrium in which the types separate after the first success. In the short-run, there may also exist a pooling equilibrium in which both types always propose the same bonus, but separation may occur over time through the evolution of the agent's beliefs. However in neither of these cases can the high type guarantee separation: there is always positive probability that the agent's belief deteriorates to the point that he ceases to supply effort before the high type principal has an opportunity to prove her type.

I will consider the Perfect Bayesian Equilibria of this game. In each period the agent chooses effort to maximise his expected utility given his belief about the principal's type $p_t$ and the promised bonus $b_t$. The agent's strategy will be a function $e : [0,1] \times \mathbb{R}_0^+ \to \{0,1\}$, that maps from the agent's belief and the promised bonus to his binary choice of effort. In each period $t$ the agent's beliefs will be determined by Bayes' rule as a function of his prior belief $p_{t-1}$, and his knowledge of output, effort and the bonus in the previous period. At the start of each period the principal offers a non-negative bonus $b_t \in \mathbb{R}_0^+$, where $b_t(h^t, \theta_i)$ may depend on the entire history of the game, and the principal's type. The principal's strategy will also specify the bonus to be paid, $\beta : \mathbb{R}_0^+ \times \{0,1\} \times \{\theta_L, \theta_H\} \to \mathbb{R}_0^+$, where the principal's choice of payment will depend upon the bonus promised, the level of output realised and her type.

I will also place the following restriction on the agent's beliefs:

**Assumption 1.** *The agent's beliefs about the principal's type are independent of the fixed* *salary s*.

This assumption rules out equilibria in which the agent updates his beliefs about the principal's type on the basis of the salary offered. For example, this would preclude an equilibrium in which the high type sought to separate by offering a contract with a large fixed component at the start of the game, or by offering efficiency wages. This implies that any separation of the two types will come through the bonus payments promised, and whether or not the bonuses are paid. This assumption also ensures that the principal does in fact have the bargaining power: I rule out situations in which the agent can obtain all the surplus from the relationship by threatening to assume that the principal is a low type if she pays any other salary but that demanded by the agent.

*Equilibrium when the principal's type is common knowledge*

Before turning to the case in which the principal's type is private information, we can find the optimal relational contract in the game in which the probability $\pi$ of high effort generating high output is common knowledge. Following Levin (2003), any payoffs that can be obtained through a relational contract in this setting can achieved by using a stationary contract. The contract will take the form of a base salary $s$, and a non-contractible bonus $b$ to be paid if output is high: $w = s + b\mathbb{I}\{y = Y_H\}$. I will look for an equilibrium in which the agent supplies effort in the first period, and then in any time period $t$, if for all $s < t$ if $y_s = Y_L$ then $b_s = b$, $e_t = 1$. Otherwise, if the principal has ever reneged on the bonus then the agent will supply zero effort in all future periods. The bonus and salary must be such that: (1) the agent is willing to participate, (2) the agent will supply costly effort, (3) the principal will pay the promised bonus when $y_t = Y_H$.

In this case, these conditions become:

1. $\frac{1}{1-\delta}[s + \pi b - c] \geq \frac{U_0}{1-\delta}$
2. $s + \pi b - c + \frac{\delta}{1-\delta}[s + \pi b - c] \geq s + \frac{\delta}{1-\delta} s$
3. $-b + \frac{\delta}{1-\delta}[\pi(Y_H - b) + (1-\pi)Y_L - s] \geq \frac{\delta}{1-\delta}(Y_L - s)$

The two incentive constraints can be satisfied whenever there exists $\delta$ such that

$$\frac{c}{\pi} \leq b^*(\pi) \leq \frac{\delta\pi(Y_H - Y_L)}{1 - \delta + \delta\pi} \tag{3.1}$$

and then the salary can be set to ensure participation: $s = U_0$. Note that the bonus paid is decreasing in $\pi$, but the principal's willingness to pay is increasing in $\pi$. Therefore a better type can ensure cooperation for a larger range of discount factors, and moreover can do so by promising a smaller bonus. For the majority of this paper I will focus on parameter values for which equation (3.1) can be satisfied for $\pi = \pi_H$, but not for $\pi = \pi_L$. That is, I will assume that were her type known, a high-type principal could obtain effort from the agent through offering a relational contract that promises a bonus $b_H = \frac{\delta\pi_H(Y_H - Y_L)}{1-\delta+\delta\pi}$ following high output, but a low-type principal would be unable to obtain any effort from the agent, since she would be unable to credibly offer the necessary bonus payment.

### 3.3.1 Equilibrium when the principal's type is private information

In this section I consider the full model in which the principal's type is private information. As noted above, the lower type has two possible motivations for imitating the higher type: firstly, for certain values of the discount factor only the high type will be able to obtain high effort from the agent; moreover even if the low type can obtain effort, the high type is able to do so through a lower bonus payment. Given that the agent can observe output and must thereby update his beliefs about the principal's type, there cannot be a fully-pooling equilibrium in the long-run. However rather than waiting for the agent to update his beliefs, the principal may prefer to signal her type in order to induce separation. I focus on the case in which only the high type of principal is able to elicit effort in a separating equilibrium, and show that the high type cannot separate from the low type through the contract offered, but can do so by paying a larger bonus after the first success. However if the agent's early attempts are unsuccessful then the agent will update his belief about the principal's type, and may cease to supply any effort, meaning that the types can never separate.

*Evolution of the agent's beliefs*

Before turning to the details of the equilibria, it will be useful to specify how the agent's beliefs will evolve over time based on his observation of output. Let $p_t$ denote the probability that the agent attaches to the principal being a high type in period $t$, where $p_1 = p$. Given that the principal's type determines the probability that high effort generates high output, the agent will update his belief based on output in each period. His beliefs will evolve in the following manner: if the agent supplied high effort and high output was realised, then

$$p_{t+1}(e_t = 1, y_t = Y_L) = \frac{p_t\pi_H}{p_t\pi_H + (1-p_t)\pi_L},$$

whereas if output is low

$$p_{t+1}(e_t = 1, y_t = Y_L) = \frac{p_t(1-\pi_H)}{p_t(1-\pi_H) + (1-p_t)(1-\pi_L)}.$$

Otherwise, if the agent does not supply effort then output cannot be high for either type, and so the agent will not update his beliefs: $p_{t+1}(e = 0) = p_t$.

Assume that at the start of period $t+1$, the agent has supplied high effort in all prior periods, and that high output has been generated $n$ times. The agent's belief $p_{t+1}$ is:

$$p_{t+1}(n) = \frac{p\pi_H^n(1-\pi_H)^{t-n}}{p\pi_H^n(1-\pi_H)^{t-n} + (1-p)\pi_L^n(1-\pi_L)^{t-n}}$$

We can also consider the expected value of this belief, conditional on the principal being a high or low type:

$$E[p_{t+1}|\pi = \pi_i] = \sum_{n=0}^{t} \binom{t}{n} \pi_i^n (1-\pi_i)^{t-n} \left[ \frac{p\pi_H^n(1-\pi_H)^{t-n}}{p\pi_H^n(1-\pi_H)^{t-n} + (1-p)\pi_L^n(1-\pi_L)^{t-n}} \right]$$

where $i = L,\ H$. It follows that if the agent supplies effort in every period, then $t \to \infty$, $E[p_{t+1}|\pi_H] \to 1$, whereas $E[p_{t+1}|\pi_L] \to 0$.

**Proposition 2.** *There is no long-run pooling equilibrium in this game in which positive effort is supplied.*

This result states that as the agent acquires more information, he must be able to distinguish the two types. That is, if $e_t = 1$ for all $t$, then as $t \to \infty$, $\Pr[p_t > 1 - \varepsilon | \theta = \theta_H] \to 1$ and $\Pr[p_t < \eta | \theta = \theta_L] \to 1$ for all $\varepsilon, \eta > 0$. This is an immediate consequence of the agent's updating of his beliefs. We know that, if the agent were to supply effort to both types in every period $t$, then as $t \to \infty$ then $p_t \to 1$ if $\theta = \theta_H$. In order for the two types to pool in the long run, it must be the case that the agent's expected beliefs about the principal's type are the same for both the high and low types. However given that the high type will have a higher rate of success conditional on the agent supplying high effort, and since the agent can observe output, the agent must update his beliefs.

Whilst there can never be a long-run pooling equilibrium in which positive effort is supplied, if the agent ceases to supply effort then he will cease to update his beliefs. In this case we could see an equilibrium in which both types promise the same bonus, but since effort is never supplied there is no chance of output being high, and therefore no chance of the bonus being paid. Moreover if the low type is unable to obtain effort when her type is known, then there must be some critical value $\hat{p}$ such that if his belief drops below this level then the agent will cease to supply effort, and therefore will no longer update his beliefs.

### 3.3.2 Separating equilibrium in the game with private information

Whilst there cannot be a fully pooling equilibrium with positive effort in the long-run, since the agent must update his beliefs, I now consider whether the high type can improve on this outcome by offering a contract or making a bonus payment that induces separation earlier in the game. I will focus on the case in which the low type is unable to elicit effort when her type is known, but the high type can, and show that in this case there is no equilibrium that induces separation at the very beginning of the game. I then show that by paying a larger bonus the high type can induce separation after high output is realised for the first time. However, there will be a positive probability that the agent's belief deteriorates to a point at which he is no longer willing to supply effort before the high type has a chance to prove her type in this way. In this case there will be no further opportunity for separation, and so even a high-type principal may find herself trapped in an equilibrium in which she cannot obtain any effort from the agent.

Assume that $\delta$ is such that equation (3.1) cannot be satisfied for $\pi = \pi_L$, but is satisfied for $\pi = \pi_H$. That is, in a separating equilibrium the low-type principal would be unable to obtain any effort from the agent. Therefore the low type will always have an incentive to imitate the high type at the start of the game: even if the high type is able to separate by paying a larger bonus following the first success, achieving high output once before reverting to the zero-effort equilibrium is strictly better than receiving zero effort throughout.

**Proposition 3.** *If $\pi_L$ is such that the low type cannot obtain effort when her type is known, then an equilibrium in which the two types separate through the contract offered and the high type obtains high effort with positive probability cannot exist.*

*Proof.* Assume that such an equilibrium did exist. That is, the high type can offer some contract $(b_t, s_t)$ that generates belief $\Pr[\theta = \theta_H | \{s_t, b_t\}] = 1$, and which induces the agent to supply effort with positive probability. We know that the low type cannot elicit effort from the worker in this case, and so her payoff in a separating equilibrium will be $Y_L - U_0$ in every period. Now consider the deviation in which the low type offers the same bonus payment as the high type. Given assumption 1, the agent's belief will not be influenced by the fixed salary offered and so the low type does not need to match the high type's fixed salary $s_t$; this also implies that the high type has no reason to offer a salary greater than that needed to ensure the agent participates. By offering this bonus, the low type can induce the agent to supply effort. If output is low, then no bonus payment is required, and so the agent will not update his beliefs if it is not paid. Otherwise, if output is high, the low type can renege on the bonus. If she does so, the agent will cease to supply effort, leaving the low type with the same payoff as she would have obtained by offering the low-type contract at the start of the game. Therefore the low-type must be made better off by deviating and imitating the high type in order to achieve high output once, without making any bonus payments. □

Although the types cannot separate through the contract offered, given that the low type can never be made worse off than she will be through separating at the beginning of the game and never obtaining any effort, the types may be able to separate on the equilibrium path if the high type can pay a bonus on which the low type would always prefer to renege. We can therefore consider an equilibrium of the following form: at the start of each period the high type of principal will offer a bonus $\hat{b}(p_t)$ conditional on success. This bonus will be sufficiently large that the low type will never pay it, but the high type will. Therefore, following output $y_t = Y_H$, the high type will pay the bonus and the low type renege, leading the agent to assign $\Pr[\theta = \theta_H | y = Y_H, b_t = \hat{b}(p_t)] = 1$, and $\Pr[\theta = \theta_H | y = Y_H, b_t \neq \hat{b}(p_t)] = 0$.

The following conditions are necessary for the bonus payment $\hat{b}(p_t)$ to support separation at the end of period $t$. Firstly, it must be such that the low type prefers to renege on the bonus and take her outside option of zero effort in all future periods. That is, even if after paying the bonus the agent will attach probability 1 to the principal being the high type, the low type must still prefer to renege:

$$-\hat{b} + \frac{\delta}{1-\delta}\left[\pi_L(Y_H - b_H) + (1-\pi_L)Y_L - s\right] \leq \frac{\delta}{1-\delta}(Y_L - s) \tag{3.2}$$

where $b_H$ is the bonus that will induce the agent to supply high effort when $p_t = 1$, i.e., $b_H = \frac{c}{\pi_H}$.

Secondly, the bonus must be incentive compatible for the high type, given that paying it will ensure that the agent believes that the principal is the high type with probability 1:

$$-\hat{b} + \frac{\delta}{1-\delta}\left[\pi_H(Y_H - b_H) + (1-\pi_H)Y_L - s\right] \geq \frac{\delta}{1-\delta}(Y_L - s) \tag{3.3}$$

Thirdly, the agent must be willing to supply high effort, given the promised bonus and her belief about the probability that it will be paid. In each period $t$, the agent will attach probability $p_t$ to the principal being a high type, and the probability of success conditional on the principal being this type is $\pi_H$. The agent's expected payoff in this case is:

$$\begin{aligned} U(p_t, \hat{b}) = s - c + p_t\pi_H[\hat{b} + \delta U(p_{t+1} = 1)] + (1-p_t)\pi_L\delta U(p_{t+1} = 0) \\ +[p_t(1-\pi_H) + (1-p_t)(1-\pi_L)]\delta U(p_{t+1}(y_t = Y_L)) \end{aligned} \tag{3.4}$$

where $U(p_{t+1} = 1) = \frac{1}{1-\delta}U_0$ is the agent's continuation value if he is paid the bonus and so attaches probability 1 to the principal being the high type; $U(p_{t+1} = 0) = \frac{1}{1-\delta}U_0$ is the continuation value when he is convinced the principal is the low type;[1] and $U(p_{t+1}(y_t = Y_L))$ is the agent's continuation value if output is low. In this last case, since output is low, neither type will pay the bonus and so the agent cannot be certain about the principal's type, but nonetheless he will update his beliefs as shown above.

The agent will be willing to supply effort if

$$U(p_t, \hat{b}) \geq \frac{1}{1-\delta}U_0 \tag{3.5}$$

In order to pin down the agent's continuation value, we must consider two cases. Since the principal has the bargaining power, she will set the minimum bonus necessary given that this bonus must both induce separation and be sufficiently large to motivate the agent. For the latter, the size of the bonus will therefore depend on the agent's belief about the likelihood that the bonus will be paid: $\Pr[b_t = \hat{b}] = p_t\pi_H$.[2] Therefore the required bonus $\hat{b}$ will depend on which of these constraints binds.

First consider the case in which the agent's incentive constraint binds in period $t$, given

[1] Since the principal has the bargaining power, the agent can be held to his outside option in either case

[2] Note that in this equilibrium neither type will ever pay a smaller bonus: on observing a smaller bonus the agent will update her beliefs to $p(b_t \neq \hat{b}) = 0$, and will therefore provide no effort in the future. Therefore if the principal is to renege on the bonus it is optimal to do so by setting $b_t = 0$.

some bonus $\hat{b}(p_t)$. Assume that the agent provided high effort, but output was low. In the subsequent period the principal must promise a higher bonus to induce the agent to supply effort, in which case the incentive constraint will once again hold with equality, and so

$$U(p_{t+1}(y_t = Y_L)) = \frac{1}{1-\delta}U_0.$$

However as $p_t$ decreases, the bonus required to induce the agent to supply effort increases. This bonus may eventually reach a level at which it is no longer incentive compatible for the high type. In this case the agent will cease to provide any effort, and will therefore never expect to receive a bonus payment, and so

$$U(p_{t+1}(y_t = Y_L)) = \frac{1}{1-\delta}U_0.$$

This expression for the agent's continuation value relies on the fact that the principal can hold the agent to his outside option in every period. This assumption can be justified from the fact that the principal has the option of hiring another agent in the future, who will have observed the history of output and the transfers paid. This implies that once the principal has succeeded in conveying her type, it will also be known to all other potential agents. She would therefore have the option of offering a different agent the minimum bonus $b_H = c/\pi_H$ required to induce the agent to supply effort. This restriction rules out an equilibrium in which the principal obtains effort from the agent by not only promising a higher bonus after the first success, but by also promising higher bonuses in the future, i.e. by promising to transfer surplus to the worker. If the principal can obtain effort from a different agent through a contract that allocates all surplus to the principal then any promises to pay a larger bonus in the future, after the types have separated, will not be credible.

This implies that in the case in which the agent's effort constraint binds, equation (3.4) reduces to

$$-c + p_t\pi_H\hat{b} + \frac{\delta}{1-\delta}U_0 \geq \frac{\delta}{1-\delta}U_0$$

which implies that $\hat{b} \geq \frac{c}{p_t\pi_H}$. Otherwise, it will be the low type's reneging constraint that binds, and so the high type can induce separation by paying the minimum bonus such that the low type will prefer to renege.

**Proposition 4.** *There exists an equilibrium in which in every period $t$ if*

$$p_t \geq \bar{p} = \frac{c(1-\delta)}{\pi_H\delta[\pi_H(Y_H - Y_L) - c]} \tag{3.6}$$

*then the agent supplies effort $e_t = 1$ and the two types separate if high output is produced.*

This result follows immediately from the preceding analysis. The principal can induce separation following high output in period $t$ by promising a bonus

$$\hat{b}(p_t) = \max\left\{\frac{\delta}{1-\delta}\left[\pi_L(Y_H - Y_L) - c\frac{\pi_L}{\pi_H}\right], \frac{c}{p_t \pi_H}\right\}. \tag{3.7}$$

We only need to verify that the high type's promise to pay the bonus $\hat{b}$ is credible. We know that if

$$\hat{b}(p_t) = \frac{\delta}{1-\delta}\left[\pi_L(Y_H - Y_L) - c\frac{\pi_L}{\pi_H}\right]$$

then the high type must be willing to pay the bonus; in this case the low type is indifferent, but the continuation value for the high type is strictly greater than for the low type. Otherwise, if the agent's incentive constraint binds, then the bonus required to obtain effort must be credible for the high type, which requires that:

$$\hat{b}(p_t) \leq \frac{c}{p_t \pi_H} \leq \frac{\delta}{1-\delta}[\pi_H(Y_H - Y_L) - c], \tag{3.8}$$

which is equivalent to the expression above.

I now turn to the dynamics of this equilibrium. Since the bonus payment necessary to induce separation may depend on the agent's belief, if the high type is unsuccessful at the beginning of the game then the amount she will need to pay the agent following the first success may increase. Figure (1) gives an example of how the promised bonus may vary over time. At the start of the game, the binding constraint on the bonus is that the low type will not imitate the high type. In this case the required bonus does not depend on $p_t$, but must exceed the low type's expected gain from paying the bonus and successfully imitating the high type. For $p_t$ sufficiently high, the agent will attach a high enough probability $p_t \pi_H$ to receiving the bonus that he will be willing to supply effort. However following a run of low output the agent's belief will deteriorate. Eventually the agent's effort constraint will bind, and so the principal will need to increase the promised bonus payment in order to ensure that the agent still supplies effort. Since the necessary bonus payment is now decreasing in $p_t$, with every successive failure the promised bonus must increase further. If the required bonus payment reaches a point at which it is no longer credible, given the principal's continuation value, then the agent will cease to supply effort. This will occur when $p_t < \bar{p}$, i.e. when the agent's belief falls below the threshold defined in equation (3.6). At this point the agent will no longer supply effort, and so the promised bonus is irrelevant. Figure (1) shows the point at which this may occur, assuming a history in which $y_t = Y_L$ for all $t$. If, at any time $t$ output had been high, the high type would have paid the promised bonus while the low type would renege. This would be sufficient to separate the two types, after which the high-type principal could revert to paying the optimal bonus $b_H$, indicated by crosses in the figure.

We can see from this analysis that not only is the high-type principal unable to separate through the contract offered at the start of the game, but that she may not be able to

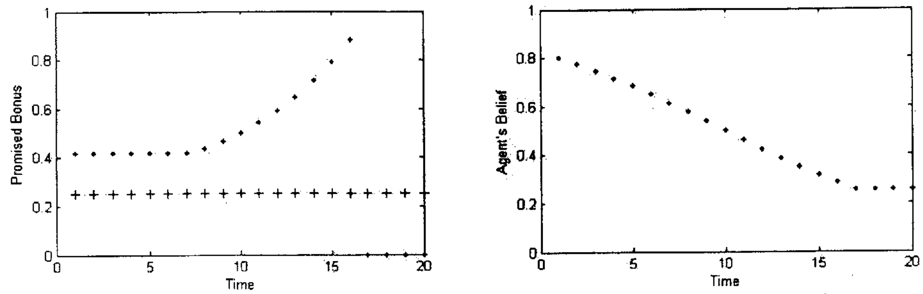


Figure 3-1: Evolution of the bonus payment and the agent's beliefs after successive failures

separate at all if there is a run of low output at the start of the game. This example also suggests that firms that have recently suffered a run of poor results may need to do a lot to restore worker confidence, and in particular will be prepared to pay a much larger bonus after a period of low output than after a period of greater success, even though we may imagine that the firm's balance sheet is healthier in the latter case.

### 3.3.3 Equilibrium in which both types pool in the short-run

I now consider whether the high type will want to offer the bonus that induces separation, given that she may be able to do better by initially pooling with the low type, optimistic that eventually the agent's beliefs will update, allowing her to pay a bonus approximately equal to that which she would pay in the game with full information. In this section I show that whilst a long-run pooling equilibrium is not possible, at the start of the game both types may be willing to promise, and pay, the same bonus. In this case the necessary bonus will depend on the agent's beliefs, which in turn depend on the history of output. I show that if the principal is in fact a high type then in general the long-run bonus that she will pay will approach the same level as in the separating equilibrium; however there is a risk that if output is low for successive periods at the start of the game then the agent will cease to supply effort, and therefore no longer update his beliefs. In contrast, if the principal is a low type then the agent's belief will eventually decrease to a point at which the necessary bonus is only credible for the high-type, and in this case separation must occur following high output.

We know that the optimal bonus defined in equation (3.7) will induce separation after high output is produced, provided that $p_t$ is such that the agent is willing to supply high effort. However we could also consider an equilibrium in which the following bonus is offered in period $t$:

$$\tilde{b} = \frac{c}{p_t\pi_H + (1 - p_t)\pi_L} \tag{3.9}$$

This bonus will be sufficient to induce effort from the agent if he believes that both types of principal will pay the bonus, given their expected future surplus. I now find conditions under which such an equilibrium will exist at the start of the game, and consider when the high-type will prefer to initially pool with the low type rather than attempting to induce separation. Since the low-type of principal cannot obtain effort in a separating equilibrium, we know that the bonus required to obtain effort from the agent $b_L \geq c/\pi_L$ cannot be credibly promised by the low type, and so

$$\frac{c}{\pi_L} > \frac{\delta\pi(Y_H - Y_L)}{1 - \delta + \delta\pi_L}.$$

Given that $\pi_H > \pi_L$, for any $p_t > 0$, $\tilde{b} < b_L$. Therefore it may be the case that the low type can commit to this bonus, given his continuation value. I will consider three cases: (i) the case in which the low-type can never credibly promise this bonus, whatever the agent's belief after the bonus is paid; (ii) that in which the low type may be willing to pay the bonus, if the agent's belief after receiving the bonus is sufficiently high; and (iii) that in which the low type strictly prefers to pay this bonus than to renege. In the first case, the equilibrium defined above, which induces separation after the first success, will be the unique equilibrium in which positive effort is supplied. In the second case, we will have a semi-separating equilibrium in which the low-type pays the bonus with some probability; this probability will be defined so as to set the agent's belief following payment, and hence the bonus in subsequent periods, such that the expected continuation surplus for the low type makes her indifferent. In the third case we will see pooling at the start of the game, but as $p_t$ decreases the game will move to either the semi-separating or fully separating phase. I will consider each of these possibilities in turn.

In the first case, the low type is unwilling to pay the bonus $\tilde{b}_t = \frac{c}{p_t\pi_H + (1-p_t)\pi_L}$ regardless of the agent's belief after receiving the payment. In particular, this implies that the low type would renege on the bonus even if paying it induced the agent to believe that $p_{t+1}(y_t = y_H, b_t = \tilde{b}_t) = 1$, i.e., that the principal must be the high type. This implies that

$$\tilde{b}_t = \frac{c}{p_t\pi_H + (1 - p_t)\pi_L} > \frac{\delta}{1 - \delta}\left[\pi_L(Y_H - Y_L) - c\frac{\pi_L}{\pi_H} -\right]$$

and so the pooling bonus is greater than the bonus required to induce separation. Therefore the only bonus that can induce the agent to supply effort is one that only the high-type would pay. As long as the high type is in fact willing to pay this bonus $b = c/p_t\pi_H$ (since the agent knows the low type will not pay, regardless of output), this will be the unique equilibrium of the game.

**Proposition 5.** *If*

$$\frac{c}{p\pi_H + (1 - p)\pi_L} > \frac{\delta}{1 - \delta}\left[\pi_L(Y_H - Y_L) - c\right]$$

*then the equilibrium that induces separation after high output is first realised is the unique* *equilibrium for which* $e_1 > 0$.

I now consider the second case, in which there may be a semi-separating equilibrium in which the low type pays the promised bonus with probability $\phi$. This will occur if

$$-\tilde{b}_t + \delta E[\Pi(\theta_L, p_{t+1}(y_t = Y_L, b_t = \tilde{b}_t))] < \frac{\delta}{1-\delta}(Y_L - U_0)$$

for $p_{t+1} = \frac{p_t \pi_H}{p_t \pi_H + (1-p_t)\pi_L}$, but

$$-\frac{c}{p_t \pi_H} + \delta E[\Pi(\theta_L, t_{p+1} = 1)] > \frac{\delta}{1-\delta}(Y_L - U_0).$$

This means that the bonus payment is such that if both types pay, then the low type's expected continuation surplus, given the way in which she expects the agent's beliefs to evolve, is insufficient to induce the low type to pay the bonus. If this were the case, then following receipt of the bonus the agent should attach probability 1 to the principal being the high type, and accept a smaller bonus $c/\pi_H$ in the future. However the low-type would then be willing to pay the bonus, given the future surplus she could obtain by successfully imitating the high type. This implies that there will exist a semi-separating equilibrium in each period in which the low-type reneges on the bonus with probability $\phi$, where:

$$\begin{aligned}
-b_t(p_t, \phi) + \delta E[\Pi(\theta_L, p_{t+1}(y_t = Y_H, b_t = b_t(p_t, \phi), \phi)] &= \frac{\delta}{1-\delta}(Y_L - U_0) \\
p_{t+1}(y_t = Y_H, b_t = b_t(p_t, \phi), \phi) &= \frac{p_t \pi_H}{p_t \pi_H + (1-p_t)\phi \pi_L} \\
b_t(p_t, \phi) &= \frac{c}{p_t \pi_H + (1-p_t)\phi \pi_L}
\end{aligned}$$

The first of these conditions ensures that the low type is indifferent between paying the bonus and reneging, given her continuation value $E[\Pi(\theta_L, p_{t+1}(y_t = Y_H, b_t = b_t(p_t), \phi)]$ which is a function of her type and the agent's updated belief. The second expression defines this updated belief as a function of $\phi$. The third expression ensures that the promised bonus $b_t(p_t, \phi)$ is sufficient to induce the agent to supply effort, given that he knows there is a positive probability that the principal may be a low type and renege.

The third case is that in which the low-type strictly prefers to pay the minimum bonus $\tilde{b}$ required to obtain effort from the agent to reneging and obtaining zero effort for the rest of the game. This requires that

$$-\tilde{b}_t + \delta E[\Pi(\theta_L, p_{p+1}(y_t = Y_L, b_t = \tilde{b}_t))] \geq \frac{\delta}{1-\delta}(Y_L - U_0),$$

meaning that this equilibrium can only exist for strictly greater values of $p_t$ than the semi-separating or separating equilibria.

We can now consider how the low-type's continuation surplus will evolve, and how that will determine which of these three outcomes may occur in any period. In the first case the agent knows that the low type will never pay the bonus $c/p_t\pi_H$, and so this is the same as in the equilibrium in which separation is induced after the first success. Following a success therefore, if the bonus is paid then the agent will attach probability 1 to the principal being a high type; this type can subsequently pay the bonus $b_H$, and receive effort $e_t = 1$ in every period. Following a failure, $p_t$ will deteriorate, implying that the bonus must increase in subsequent periods in order to persuade the agent to supply effort. Eventually we will reach a point at which $p_t < \bar{p}$, and so the bonus necessary for the agent to supply effort is such that even the high type would prefer to renege; anticipating this, the agent will cease to supply any effort.

This equilibrium in which separation is induced after the first success is absorbing, in the sense that once $p_t$ is such that this is the unique outcome in period $t$, the game will never revert to one of the other equilibria in which the low type pays the bonus with positive probability. This is because if output is high full separation will occur, but if output is low the agent's beliefs can only decrease further, implying that the bonus will increase and so will still not be credible for the low type.

In contrast, if $p_t$ is such that either the semi-separating or pooling outcome is possible in period $t$, then we may see the parties move between these phases. For example if $p_t$ is high, and therefore $\tilde{b}$ smaller, the bonus may be credible for both types. Following a success, therefore, both types will pay the bonus and so $p_{t+1} > p_t$, reducing the requisite bonus further, increasing the continuation value for the low type, and ensuring that this pooling outcome will exist in the following period. On the other hand, if output is low then $p_{t+1} < p_t$: in this case we may move from the phase in which the low type always pays the bonus to that in which she randomises, or to the separating phase in which the low type will never pay the bonus. Moreover, we know from Lemma 1 that as $t \to \infty$, if the principal is in fact a low type, then the agent's belief must tend towards zero, implying that it must eventually fall below $\bar{p}$, and so in the long-run the low-type will be unable to obtain any effort from the agent.

Figures (2) and (3) show two possible histories, in which the principal is a high type and a low type respectively. In the first figure, we can see the evolution of the bonus payment required to elicit costly effort from the agent as his beliefs evolve. Over time, the agent's belief that the principal is a high type approaches 1, and so the high-type principal can obtain effort by promising approximately the same bonus as she would in a separating equilibrium. Moreover, by this point low output in a particular period has little effect on the agent's beliefs, and so the bonus will remain fairly constant. We can contrast this outcome when the principal is a high type with that in which the principal is a low type. In this example, the low type is able to obtain effort from the agent at the start of the game. That is, early successes mean that the agent's belief about the principal's type is quite high, and so the low type can obtain effort by promising a smaller bonus, one which she is willing to pay. However following a run of low output the agent's belief deteriorates, and so in order

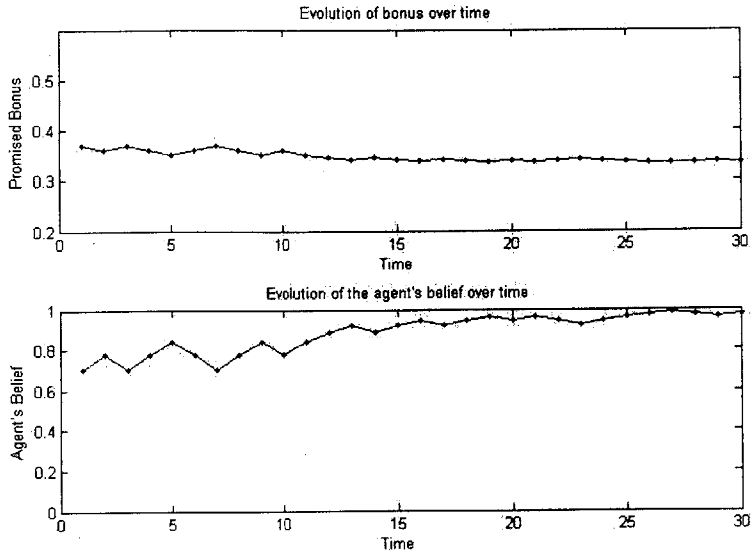


Figure 3-2: Example: equilibrium with initial pooling when the principal is a high type

to obtain high effort the bonus must increase. In particular, the bonus will increase rapidly once it has reached a level at which the low type will renege. In this case if high output is achieved the two types will separate: the low type will renege and so the agent will update his belief to $p_t(y_t = Y_H, b_t = 0) = 0$.

We can also see from this example that a low type is more likely to completely separate than a high type. In this case full separation will only occur if in some period $t$ the bonus payment required to induce effort is such that the high type will pay the bonus following high output, but the low type will renege. However the bonus payment will only reach this level when $p_t$ is small, that is, if output has frequently been low. Given the differences in success probabilities $\pi_H$ and $\pi_L$, such a history is more likely if the agent is in fact a low type. In contrast, if the principal is a high type then we will expect to see more early successes, which will increase the agent's belief $p_t$, reducing the necessary bonus and making full separation (i.e., the agent assigning belief $p_t = 1$ to the principal being a high type) less likely. In this case early successes, rewarded with more moderate bonus payments, are sufficient to build tha agent's confidence in the principal's type, and so she will not need to make stronger promises the agent in order to obtain this trust and secure the future of the relationship.

### 3.3.4 Optimal equilibrium for the high-type

Having considered the two possible equilibria in which the agent supplies effort at the start of the game, I now turn to the question of which is optimal. The high type faces a trade-off here: if she initially pools with the low type then in expectation the agent will update his

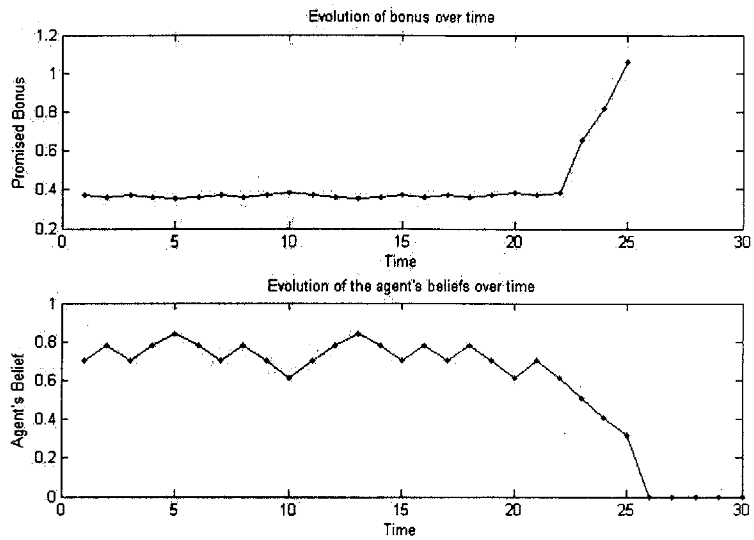


Figure 3-3: Example: equilibrium with initial pooling when the principal is a low type

beliefs in such a way that the long-run equilibrium will be the same as following separation. However this learning may take a long time. In contrast, if she promises a larger bonus following the first success then the high type will expect separation to occur more quickly, but she must incur this greater cost. In either case there is a risk that the agent's belief may deteriorate to a level at which he no longer supplies effort; this is a greater risk in the pooling equilibrium, since in the separating equilibrium only one success is required to ensure this outcome will not occur.

In this section I show that the high type will prefer to offer a larger bonus after the first success in order to distinguish herself if learning is otherwise likely to be slow, or if the bonus payments at the start of the game in the pooling equilibrium will be large. This implies that the separating equilibrium is preferred if $p$ is small, or if the difference between $\pi_H$ and $\pi_L$ is small. We can see these results by comparing the expected surplus to a high type principal from the pooling equilibrium and the separating equilibrium. This expected surplus must be computed numerically, and figure (4) shows how it varies with $p$, the probability that the principal is in fact the high type, and $\pi_H - \pi_L$, the difference in success rates.

As noted in Lemma (2), there are parameter values for which the separating equilibrium is the unique equilibrium in which effort is supplied, since the bonus required to elicit effort from the agent is such that the low type will always renege. We also saw above that even if the relationship begins in a pooling equilibrium in which both types pay the promised bonus, after certain histories the relationship may move into this separating phase, even if the principal is in fact a high type. Moreover, the principal herself may wish to switch initially attempting to pool with the low type, to promising a larger bonus payment that would induce separation following the next instance of high output. We can see this by

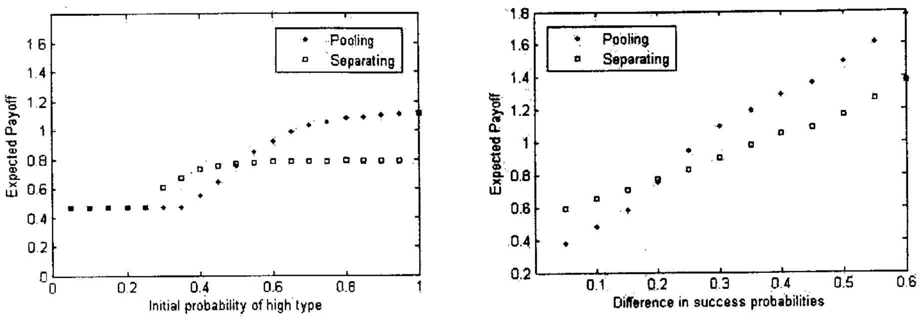


Figure 3-4: Comparing the expected payoffs from pooling vs. inducing separation

comparing the expected payoffs from the two equilibria as a function of $p$. If output is low at the start of the game then the agent's belief deteriorates: this both increases the bonus that must be paid, and increases the risk of the agent withdrawing effort altogether. The high type might therefore want to increase the promised bonus beyond that required to obtain effort from the agent: for $p$ small, the increase in the bonus paid following the first success will be small, and a single realisation of high output is sufficient to ensure that the high type receives high effort for the remainder of the game.

In contrast, the principal will never abandon a strategy of attempting to separate in favour of pooling with the lower type: pooling will only be preferred if $p_t$ has increased compared to $p_1 = p$. However this can only occur following a success, in which case the high type will have already succeeded in fully separating by paying the larger bonus.

These results suggest that if the agent's belief $p$ at the start of the game is high then we would expect to see both types promising and paying the same bonus payments. During a run of high performance, the agent's trust in the firm will increase and there will be no need for bonus payments to increase. However, following a run of poor performance the agent's beliefs will decrease. Even if the types continue to pool, we would therefore expect to see higher bonuses being paid as the firm starts to recover and output increases. This effect will be amplified if $p_t$ has dropped to the level at which the high type now prefers to separate at the earliest opportunity, rather than risk being trapped in the long-run equilibrium in which no effort is supplied. In this case some firms may pay much larger bonuses after a period of poor performance, whilst others will renege and accept that the agent will punish this by ceasing to supply effort. This result suggests that we may see greater heterogeneity in bonus payments, and more breakdowns in firm-employee relationships as companies start to recover from a period of economic turmoil.

## 3.4 Matching the Market

In this section I extend the model to allow for the possibility that the agent may receive a signal of the output of a high-type principal. That is, the agent learns the probability that his effort would have generated high output, were the principal in fact a high type. The agent will take this information into account when updating his beliefs: in particular, were output low following a favourable signal the agent will attach less weight to the principal being a high type than were the signal unfavourable. I show that this will affect the equilibrium in two ways. Firstly, by influencing the agent's beliefs the signal will affect the dynamics of the bonus payment. In particular, if a run of low output is accompanied by signals suggesting that the high type is unlikely to produce high output then there is less need for the bonus payment to increase. However if the agent receives an indication that the high type should have produced high output then this will put more pressure on the bonus. Secondly, the signal may determine whether or not the principal wishes to pay a promised bonus payment, given that it will influence the continuation value.

We can think of this signal as an indication of market conditions, or the performance of similar firms. In this case the agent will compare the performance of the principal against the performance of those he regards as her competitors. If similar firms appear to be doing badly, that it, if the signal is bad, then the agent is less inclined to consider the principal to be a bad type when output is low. However if the signal is high, and so other firms appear to be doing well, then the agent will be quicker to update his beliefs and potentially to withdraw effort altogether. This indicator of market performance also affects the firm's incentive to pay a bonus: if competitors appear to be doing well then there is greater value in "matching the market" and paying the bonus; if not, then the firm has a greater incentive to renege.

Let the signal received by the agent $\sigma_t \in \{0, 1\}$ indicates whether or not the high-type will generate output $y_t = Y_H$, conditional on $e_t = 1$. Assume that the signal is correct with probability $q > 1/2$. That is, $\Pr[y_t = Y_H | \theta = \theta_H, \sigma_t = 1] = \Pr[y_t = Y_L | \theta = \theta_H, \sigma_t = 0] = q$. I will assume that the likelihood of high output being generated is independent across types, and therefore that the signal tells the agent nothing about the probability of the low-type generating high output in any period. I will assume that this signal is observed by both the principal and the agent at the same time as output is observed. Therefore the realisation of this signal will not affect the principal's choice of bonus offered, or the agent's choice of effort, but it may determine whether or not the bonus is paid, and the agent's inference about the principal's type.

The agent will now condition his beliefs both on his observation of output and on the signal received. Specifically, the agent's beliefs will now take the following form:

$$p_{t+1}(Y_H, \sigma = 1) = \frac{qp_t\pi_H}{qp_t\pi_H + 0.5(1-p_t)\pi_L}$$
$$p_{t+1}(Y_H, \sigma = 0) = \frac{(1-q)p_t\pi_H}{(1-q)p_t\pi_H + 0.5(1-p_t)\pi_L}$$
$$p_{t+1}(Y_L, \sigma = 1) = \frac{(1-q)p_t(1-\pi_H)}{(1-q)p_t(1-\pi_H) + 0.5(1-p_t)(1-\pi_L)}$$
$$p_{t+1}(Y_L, \sigma = 0) = \frac{qp_t(1-\pi_H)}{qp_t(1-\pi_H) + 0.5(1-p_t)(1-\pi_L)}$$

### 3.4.1 Inducing separation through payment of the bonus.

I first illustrate how the introduction of this signal will affect the bonus payment that must be promised in the equilibrium in which the high type endeavours to induce separation at the earliest opportunity. In this case we will see less upward pressure on the bonus payment during a run of low output if the signal is low, that is, if the agent believes that all firms are performing poorly. However the bonus will increase more rapidly, and the relationship will be more likely to break down as the agent ceases to supply effort, if the signal suggests that market conditions are good.

As in the baseline model, if the low type is unable to obtain effort when the agent knows her type, then it will not be possible for the two types to separate through the contract offered; the low type will always have an incentive to imitate the high type in order to try and obtain high output once, and then reneging on the bonus. However the high type may be able to induce output following the first success by promising a bonus payment on which the low type would always prefer to renege. As discussed above, the bonus necessary to induce separation in this case must be sufficient to induce effort from the agent, and to ensure that the low type doesn't wish to imitate. The bonus will be the minimum value of $b$ that satisfies these constraints, subject to being credible for the high type. After a run of successive failures the agent's belief will deteriorate, increasing the necessary bonus payment. However the agent's belief will now depend not only on output and his prior, but also on the signal. In particular, if the agent's signal suggests that output should be high, were the principal a high type, then low output will lead to a greater decrease in this beliefs. We can see this in figure (5) below, which shows the evolution of the agent's beliefs and the bonus promised, given the signal, if output is low in every period

Figure (5) shows how the bonus payment required to induce separation will vary with the agent's belief, which in turn depends on the signal received. The right-hand figure shows these signals: in periods in which the signal is zero, and so the agent does not expect the high type to generate high output, we see only a small decrease in the agent's beliefs, despite successive failures. When the signal is one, and so the agent expects the high type to have produced high output in that period, the belief decreases more sharply; once the agent's incentive constraint has started to bind this in turn drives up the promised bonus payment.

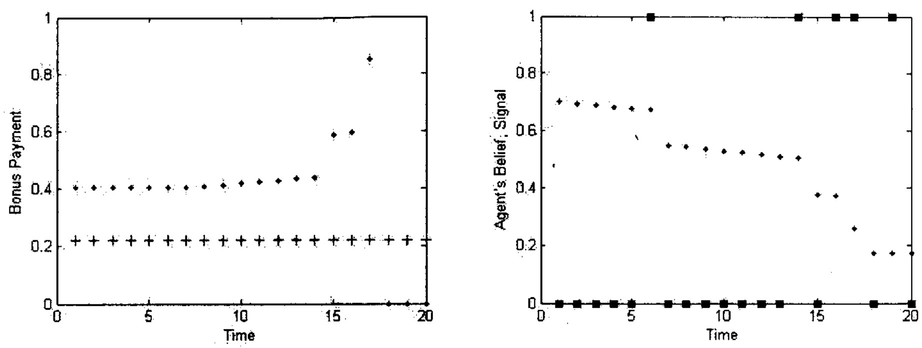


Figure 3-5: Dynamics of the bonus payment with the signal

Eventually the magnitude of the bonus, combined with the signal $\sigma_{17} = 1$ means that the agent's belief deteriorates to the point at which the bonus payment necessary to induce effort is not credible for either type.

In this case the signal only affects whether or not the bonus is paid insofar as it determines the agent's beliefs and hence the magnitude of the bonus; it does not affect the continuation value. This is because whenever this bonus is paid it must be case that the two types separate. Since output is observed, failure to pay the bonus leads to the agent ceasing to supply effort in all subsequent periods. Alternatively, if the bonus is paid then the agent attaches probability 1 to the principal being the high type and so the signal cannot affect his beliefs in the future.

We can see from the above figure that as in the baseline case there may be histories in which the high type is unable to separate and so remains trapped in an equilibrium in which the agent supplies no effort. The effects of a more informative signal are ambiguous. We know that one success is sufficient to separate the types (and thereafter the signal becomes irrelevant), so consider the case in which the principal is a high type, but suffers successive failures. If the signal is more accurate, then it is likely that the agent will receive a signal that the high type is unable to supply effort: in this case his beliefs will not deteriorate following observations of low output. This gives the principal more time to obtain the single success which can ensure the survival of the relationship. However for $q$ sufficiently large one incorrect signal suggesting that the high type should have achieved high output may be enough to lower the agent's beliefs to a point at which he ceases to supply effort altogether.

### 3.4.2 Equilibrium in which types pool in the short run

I now consider the equilibrium in which the types may pool in the short run, and so both types may promise, and pay, the same bonus in early time periods. In this case the principal's decision regarding whether or not to pay the bonus in a particular time period may depend on the signal $\sigma_t$ observed during that period, not only on how prior signals $\{\sigma_i\}_{t=0}^{t-1}$ have affected the bonus payment that must be promised in period $t$. This is because the signal in period $t$

will determine the agent's beliefs in the future, and hence the principal's continuation value. The continuation value will be higher if the signal is one than if the signal is zero, and this difference may convince the low-type principal to pay a bonus when otherwise she would have reneged. In this case the pressure to "match the market", i.e., to behave as if a high type, may convince the principal to pay the bonus; conversely, if output is low in this case then this will lead to a greater decrease in the agent's belief, and he may cease to supply effort.

Assume that at the start of period $t$ both types promise a bonus payment $b$ conditional on high output. As in the baseline game, there are three potential outcomes in each period: that in which both types pay, and so continue to pool; a semi-separating case in which the low type pays the bonus with some probability, where that probability will determine the agent's beliefs and hence the continuation value; and the case in which only the high type will pay, and so any success must induce separation. From above, we know that the conditions for an equilibrium in each period are:

$$
\begin{aligned}
b_t(p_t,\phi) + \frac{\delta}{1-\delta}(Y_L - U_0) &= \delta E[\Pi(\theta_L, p_{t+1}(Y_H, b_t = b_t(p_t,\phi), \phi, \sigma_t)], \\
p_{t+1}(Y_H, b_t = b_t(p_t,\phi), \phi, \sigma_t) &= \frac{(\sigma_t q + (1-\sigma_t)(1-q))p_t\pi_H}{(\sigma_t q + (1-\sigma_t)(1-q))p_t\pi_H + 0.5(1-p_t)\phi\pi_L}, \\
b_t(p_t,\phi) &= \frac{c}{p_t\pi_H + (1-p_t)\phi\pi_L},
\end{aligned}
$$

where $\phi$ is the probability that the low type pays the bonus: if $\phi = 0$ we have complete separation following a success, if $\phi = 1$ the types continue to pool, and for $\phi \in (0,1)$ there will be partial separation. In this case the expected continuation value for the principal will also depend on $\sigma$, since this will influence the agent's future beliefs, and thus the bonuses that must be promised in subsequent periods. By influencing the continuation surplus, the signal will therefore also influence the principal's incentive to pay the bonus.

**Proposition 6.** *There exist bonus payments $b_t(p_t)$ on which the principal will renege with positive probability if $\sigma_t = 0$, but will pay with certainty if $\sigma_t = 1$*

*Proof.* We can compare the expected continuation surplus $E[\Pi(\theta_L, p_{t+1}(Y_H, b_t = b_t(p_t,\phi), \phi, \sigma_t)]$ for $\sigma_t = 1$ and $\sigma_t = 0$; since $q > 1/2$, we must have

$$
\frac{qp_t\pi_H}{qp_t\pi_H + 0.5(1-p_t)\pi_L} > \frac{(1-q)p_t\pi_H}{(1-q)p_t\pi_H + 0.5(1-p_t)\pi_L}
$$

and so probability that the agent attaches to the principal being a high type will be greater if $\sigma = 1$ than if $\sigma = 0$. This implies that the bonus payment the agent must be promised in the following period is lower, since there is a higher probability of payment. Moreover, conditional on output and the signal, beliefs in all subsequent periods will be higher. Therefore the expected continuation value following a positive signal is higher:

$$E[\Pi(\theta_L, p_{t+1}(Y_H, b_t = b, \phi, \sigma_t = 1)] > E[\Pi(\theta_L, p_{t+1}(Y_H, b_t = b, \phi, \sigma_t = 0)].$$

This in turn increases the bonus payment that the low type is willing to pay. In particular, for certain $b$ we can have:

$$\delta E[\Pi(\theta_L, p_{t+1}(Y_H, b_t = b, \phi, \sigma_t = 0)] < b + \frac{\delta(Y_L - U_0)}{1-\delta} < \delta E[\Pi(\theta_L, p_{t+1}(Y_H, b_t = b, \phi, \sigma_t = 1)].$$

In this case the low-type would prefer to renege on the bonus payment were the signal low, but pay if the signal suggests that the high-type would be able to do so. Whilst this outcome will not occur in equilibrium (since this would imply that $p_{t+1}(Y_H, \sigma_t = 0, b) = 1 > p_{t+1}(Y_H, \sigma_t = 1, b)$ if the low type reneges with certainty), the low type will renege with positive probability following a low signal, but always pays the bonus following a high signal. This payment probability $\phi$ will be determined so as to ensure that the low type is indifferent between paying and reneging, given the continuation value. □

We can compare this outcome to that in the baseline equilibrium, or equivalently, to that when the signal conveys no information: $q = 1/2$. Assume that in this case $p_t$ and the bonus $b_t(p_t)$ are such that the low type will pay the bonus some probability, given the concomitant continuation value. If $q$ increases, then paying the bonus following a positive signal generates a higher continuation value and so must be preferred; if the signal is low then the continuation surplus decreases, and the agent will mix and pay the bonus with some probability. The more accurate the signal, the more value there is to a low-type principal in paying the the bonus and attempting to match the performance of her perceived competitors following a positive signal. On the other hand, if the signal suggests that high types are unlikely to have received high output ($\sigma = 0$), then high output may convey a negative impression to the agent.[3] In this case the principal prefers to renege on the bonus, since the agent would infer her type and cease to supply effort anyway.

In addition to affecting the principal's incentives to pay the bonus in period $t$, the signal $\sigma_t$ also influences the bonus payments that will be promised in the future, similarly to in the separating equilibrium described above. Again, this implies that even if output is low for successive periods, if the signal suggests that other firms are likely to performing badly, the agent will not punish poor performance by expecting much larger bonus payments in the future, or by ceasing to supply effort. The greatest pressure on bonus payments, and

[3]In this case, where output is independent across types the signal conveys no information about the expected output of a low type. This implies that if $\sigma_t = 0$ but $y_t = Y_H$ the weight the agent attaches to the principal being a high type decreases: that is, the positive signal from high output is outweighed by the negative signal. We could imagine this happening if unexpected success whilst other firms are performing badly is regarded as suspicious or indicative of fraudulent activities. An alternative structure could have output not be independent across types: for example we could imagine that a low type would achieve success only if a high type also would. In this case high output would reveal that a signal $\sigma_t = 0$ must be incorrect.

the greatest risk of the agent's belief deteriorating to a level at which he is no longer willing to supply effort therefore comes when the signal is positive, but output low. Therefore we would expect to see bonuses increase rapidly after a downturn as firms face greater pressure to prove their quality during the recovery. This will be accompanied by greater heterogeneity in firm performance both because those who fail to achieve high output are more likely to see the agent withdraw effort altogether, and because the higher bonus payments are more likely to induce separation as the low types renege.

## 3.5 Conclusion

This paper considered a model of relational contracting in which the principal's type determines the probability of high effort generating high output, where this type is privately known to the principal. In this setting the bonus payment required to elicit effort from the principal will be driven by the agent's beliefs, which depend on his observations of output. In particular, the promised bonus payment must increase following a run of poor performance; this is necessary to maintain the agent's confidence in the firm and ensure that he continues to supply effort. We can distinguish between the equilibrium in which the high type tries to induce separation by paying a larger bonus following the first success, and that in which both types promise, and pay, the same bonus initially until the agent's beliefs gradually update. The separating equilibrium is likely to be preferred by the high type when the belief the agent attaches to this type is lower. This implies that following a period of poor performance during which the agent's belief has declined the principal may be more inclined to try and induce separation at the earliest opportunity, even though this requires a larger bonus. Therefore we may expect to see firms paying larger bonuses after a run of poor performance, or as they emerge from a downturn, even though we would imagine that they may have less resources to do so.

The paper then considers this need to match the expectations of the agent more explicitly by augmenting the model to include a signal than may be interpreted as an indicator of market conditions or the performance of potential competitors. In this case we will see less pressure on the firm to prove its type, and hence less pressure on the bonus payments, if the signal suggests that high-type firms are unlikely to be performing well. However if the agent receives information that suggests that a high-type principal should be producing high output then he will judge underperformance more harshly, and this will exacerbate the pressure on the principal to prove her quality. The signal will affect not only the bonus that must be promised to the agent in order to maintain his confidence, but may also affect the principal's decision of whether or not to pay. In particular, the principal's continuation value from paying a bonus is greater following a positive signal, that is, when the agent expects a high type to produce high output. In this case we may see firms paying a bonus in order to match the market and keep up with those perceived as leading firms.

# Bibliography


[1] Dilip Abreu, David Pearce, and Ennio Stacchetti. "Toward a theory of discounted repeated games with imperfect monitoring". *Econometrica: Journal of the Econometric Society*, pages 1041–1063, 1990.

[2] Susan Athey and Kyle Bagwell. "Optimal collusion with private information". *RAND Journal of Economics*, pages 428–465, 2001.

[3] M.N. Baily, C. Hulten, D. Campbell, T. Bresnahan, and R.E. Caves. "Productivity dynamics in manufacturing plants". *Brookings Papers on Economic Activity. Microeconomics*, 1992:187–267, 1992.

[4] C.K. Bart, N. Bontis, and S. Taggar. "A model of the impact of mission statements on firm performance". *Management Decision*, 39(1):19–35, 2001.

[5] M.L.W. Bentley. "Optimal contracting with subjective evaluation". *The American Economic Review*, 93(1):216–240, 2003.

[6] B Douglas Bernheim and Michael D Whinston. "Multimarket contact and collusive behavior". *The RAND Journal of Economics*, pages 1–26, 1990.

[7] A. Blume and A.M. Franco. "Decentralized learning from failure". *Journal of Economic Theory*, 133(1):504–523, 2007.

[8] A. Blume, A.M. Franco, and P. Heidhues. "Dynamic coordination via organizational routines". 2011.

[9] P. Bolton, M.K. Brunnermeier, and L. Veldkamp. "Leadership, Coordination and Corporate Culture". Technical report, Working Paper (August), 2011.

[10] C. Brown and M. Reich. "When does union-management cooperation work? A look at NUMMI and GM-Van Nuys". *California Management Review*, 31(4):26–44, 1989.

[11] S.E. Cha and A.C. Edmondson. "When values backfire: Leadership, attribution, and disenchantment in a values-driven organization". *The Leadership Quarterly*, 17(1):57–78, 2006.

[12] S. Chassang. "Building routines: Learning, cooperation, and the dynamics of incomplete relational contracts". *The American Economic Review*, 100(1):448–465, 2010.

[13] W.B. Chew, T.F. Bresnahan, and K.B. Clark. "Measurement, coordination, and learning in a multiplant network". *Measures for manufacturing excellence*, pages 129–162, 1990.

[14] I.K. Cho and J. Sobel. "Strategic stability and uniqueness in signaling games". *Journal of Economic Theory*, 50(2):381–413, 1990.

[15] In-Koo Cho and David M Kreps. "Signaling Games and Stable Equilibria". *The Quarterly Journal of Economics*, 102(2):179–221, May 1987.

[16] Vincent P Crawford and Joel Sobel. "Strategic Information Transmission". *Econometrica*, 50(6):1431–51, November 1982.

[17] V.P. Crawford and H. Haller. "Learning how to cooperate: Optimal play in repeated coordination games". *Econometrica: Journal of the Econometric Society*, pages 571–595, 1990.

[18] V. Dermol. "Relationship between Mission Statement and Company Performance". In *Knowledge and Learning: Global Empowerment; Proceedings of the Management, Knowledge and Learning International Conference 2012*, pages 891–899. International School for Social and Business Studies, Celje, Slovenia, 2012.

[19] S. Desmidt and A. Heene. "Methodology in mission statement research: where are we, and were should we go? An analysis of 20 years of empirical research". 2006.

[20] T. Dewan and D.P. Myatt. "The qualities of leadership: Direction, communication, and obfuscation". *American Political Science Review*, 102(3):351–368, 2008.

[21] G. Ellison and R. Holden. "A theory of rule development". *Unpublished manu*, 2008.

[22] Florian Englmaier and Carmit Segal. "Cooperation and Conflict in Organizations". Technical report, Mimeo, 2011.

[23] J. Farrell and R. Gibbons. "Cheap talk with two audiences". *The American Economic Review*, 79(5):1214–1223, 1989.

[24] Yuk-fai Fong and Jin Li. "Relational contracts, efficiency wages, and employment dynamics". *Northwestern University, Kellogg School of Management*, 2010.

[25] L. Foster, J. Haltiwanger, and CJ Krizan. "Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s". *The Review of Economics and Statistics*, 88(4):748–758, 2006.

[26] L. Foster, J. Haltiwanger, and C. Syverson. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?". *American Economic Review*, 98(1):394–425, 2008.

[27] William Fuchs. "Contracting with Repeated Moral Hazard and Private Evaluations". *The American Economic Review*, 97(4):1432–1448, 2007.

[28] D. Fudenberg, D. Levine, and E. Maskin. "The folk theorem with imperfect public information". *Econometrica: Journal of the Econometric Society*, pages 997–1039, 1994.

[29] R. Gibbons. "Incentives between firms (and within)". *Management Science*, 51(1):2–17, 2005.

[30] R. Gibbons and R. Henderson. "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises". 2012.

[31] E.J. Green and R.H. Porter. "Noncooperative collusion under imperfect price information". *Econometrica: Journal of the Econometric Society*, pages 87–100, 1984.

[32] Marina Halac. "Relationship building". *Columbia Business School Research Paper*, (11-4), 2011.

[33] Marina Halac. "Relational contracts and the value of relationships". *The American Economic Review*, 102(2):750–779, 2012.

[34] Benjamin E Hermalin. "Toward an Economic Theory of Leadership: Leading by Example". *American Economic Review*, 88(5):1188–1206, December 1998.

[35] Benjamin E. Hermalin and Michael S. Weisbach. "Transparency and Corporate Governance". NBER Working Papers 12875, National Bureau of Economic Research, Inc, January 2007.

[36] C.T. Hsieh and P.J. Klenow. "Misallocation and manufacturing TFP in China and India". *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009.

[37] A. Iseri-Say, A. Toker, and D. Kantur. "Do popular management techniques improve performance?: Evidence from large businesses in Turkey". *Journal of Management Development*, 27(7):660–677, 2008.

[38] M. Kandori. "Introduction to repeated games with private monitoring". *Journal of Economic Theory*, 102(1):1–15, 2002.

[39] Navin Kartik. "Strategic Communication with Lying Costs". *Review of Economic Studies*, 76(4):1359–1395, 2009.

[40] Jean-Jacques Laffont and Jean Tirole. "The dynamics of incentive contracts". *Econometrica: Journal of the Econometric Society*, pages 1153–1175, 1988.

[41] L. Leuthesser and C. Kohli. "Corporate identity: the role of mission statements". *Business Horizons*, 40(3):59–66, 1997.

[42] J. Levin. "Relational incentive contracts". *The American Economic Review*, 93(3):835–857, 2003.

[43] J. Li and N. Matouschek. "Managing Conflicts in Relational Contracts". Technical report, Mimeo, Northwestern University, Evanston, IL, 2011.

[44] Bentley W. MacLeod. "Optimal Contracting with Subjective Evaluation". *American Economic Review*, 93(1):216–240, March 2003.

[45] James M Malcomson. "Relational incentive contracts with persistent private information". Technical report, 2012.

[46] M. Melitz. "The impact of trade on aggregate industry productivity and intra-industry reallocations". *Econometrica*, 71(6):1695–1725, 2003.

[47] Charles A. O'Reilly. "New United Motors Manufacturing, Inc. (NUMMI)". *Harvard Business Review*, 1998.

[48] J.A. Pearce. "The company mission as a strategic tool". *Sloan management review*, 23(3):15–24, 1982.

[49] J.A. Pearce and F. David. "Corporate mission statements: The bottom line". *The Academy of Management Executive (1987-1989)*, pages 109–115, 1987.

[50] Garey Ramey and Joel Watson. "Contractual fragility, job destruction, and business cycles". *The Quarterly Journal of Economics*, 112(3):873–911, 1997.

[51] D. Rigby and C. Gillies. "Making the most of management tools and techniques: a survey from Bain & Company". *Strategic Change*, 9(5):269–274, 2000.

[52] J.J. Rotemberg and G. Saloner. "Visionaries, managers, and strategic direction". *RAND Journal of Economics*, pages 693–716, 2000.

[53] C. Syverson. "Product substitutability and productivity dispersion". *Review of Economics and Statistics*, 86(2):534–550, 2004.

[54] C. Syverson. "What determines productivity at the micro level?". *Journal of Economic Literature*, 2011.

[55] E. Van den Steen. "Organizational beliefs and managerial vision". *Journal of Law, Economics, and Organization*, 21(1):256–283, 2005.

[56] Joel Watson. "Starting small and renegotiation". *Journal of economic Theory*, 85(1):52–90, 1999.

[57] Joel Watson. "Starting small and commitment". *Games and Economic Behavior*, 38(1):176–199, 2002.